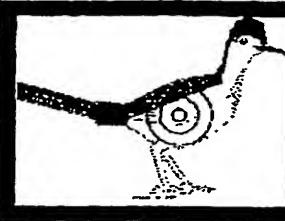


FIG. 1

1/40

File Edit View Go Communicator Bookmarks Location: http://victoria.inpharmatica.co.uk/~volker/BPD3target.html

WIPO

 Target Mining Interface



Select Your Query Sequence

• Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):

OR

• Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

• Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

Filter for the following SPECIES:

0-100%

FIG. 2A

2/40

File Edit View Go Communicator Help
 Bookmarks Location http://London-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl

2) 84 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!

Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st Iter. (PSI)	Best Iter. (PSI)	E-value (F)
J	AAA59544.1 drill through Top50BlastHits	AAA59544.1 Red.Seq.View	Not given	Homo sapiens	PRJ	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E
J	AAB24821.1 drill through Top50BlastHits	AAB24821.1 Red.Seq.View	leukocyte integrin alpha chain	Homo sapiens	PRJ	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E
J	Q99715 drill through Top50BlastHits	Q99715 Red.Seq.View	COLLAGEN ALPHA I(XII) CHAIN PRECURSOR.	Homo sapiens (Human)	PRJ	28.8%, 26% unmaskedSW	4-181, 2-174	140-318, 2321-2495	440	100% unmaskedGT	1	3	2E
J	P20701 drill through Top50BlastHits	P20701 Red.Seq.View	LEUKOCYTE ADHESION GLYCOPROTEIN LFA-1 ALPHA CHAIN PRECURSOR (LEUKOCYTE FUNCTION ASSOCIATED MOLECULE 1, ALPHA CHAIN) (CD11A) (INTEGRIN ALPHA-L).	Homo sapiens (Human)	PRJ	100%, 100% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E
J	AAC31672.1 drill through Top50BlastHits	AAC31672.1 Red.Seq.View	leukocyte function-associated molecule-1 alpha subunit	Homo sapiens	PRJ	99.5%, 99% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E
J	CAA72402.1 drill through Top50BlastHits	CAA72402.1 Red.Seq.View	collagen type XIV	Homo sapiens	PRJ	29.1%, 29% unmaskedSW	2-180, 2-180	5-185, 5-185	422	100% unmaskedGT	1	2	2E
J	AAB38702.1 drill through Top50BlastHits	AAB38702.1 Red.Seq.View	cartilage matrix protein	Homo sapiens	PRJ	31.7%, 27% unmaskedSW	4-183, 2-182	275-455, 39-223	413	100% unmaskedGT	1	2	2E
J	CAB70853.1 drill through Top50BlastHits	CAB70853.1 Red.Seq.View	hypothetical protein	Homo sapiens	PRJ	28%, 28% unmaskedSW	1-180, 1-183	437-620, 437-624	406	100% unmaskedGT	1	2	1E
J	CAA27972.1 drill through Top50BlastHits	CAA27972.1 Red.Seq.View	Not given	Homo sapiens	PRJ	20.5%, 20% unmaskedSW	3-183, 2-181	1497-1673, 1689-1873	405	100% unmaskedGT	2	3	1E
J	AAB59512.1 drill through Top50BlastHits	AAB59512.1 Red.Seq.View	Not given	Homo sapiens	PRJ	20.5%, 20% unmaskedSW	3-183, 2-181	758-934, 950-1134	405	100% unmaskedGT	2	3	1E
J	CAA07569.1 drill through Top50BlastHits	CAA07569.1 Red.Seq.View	matrix-4	Homo sapiens	PRJ	28.1%, 25% unmaskedSW	1-183, 1-183	342-528, 31-217	403	100% unmaskedGT	1	2	2E

3/40

Act/Detail	Residuante sequence display	BPD link	WWW link	Title	Organism	Dv.	%ID	Query seqn.	Target seqn.	Ahn. score	Conf.
-	Res. Seq. View	AAT71133.1	AAF71133.1	PRO2789	Homo sapiens	PRI	13.8%	unmaskedSW	09-179	1-80	122
-	Res. Seq. View	Top50BlastHits	CAB52192.1	G7c protein	Homo sapiens	PRI	9.6%	unmaskedSW	7-117	20-124	82
-	Res. Seq. View	Top50BlastHits	CAA82810.1	basic transcription factor 2, 44 kD subunit	Homo sapiens	PRI	11%	unmaskedSW	5-164	61-225	75
-	Res. Seq. View	Top50BlastHits	CAA82810.1		Homo sapiens	PRI	9.6%	unmaskedSW	7-117	318-422	82
-	Res. Seq. View	Top50BlastHits	AAD21820.1	NG37	Homo sapiens	PRI	9.6%	unmaskedSW	7-117	98-86%	unmaskedGT
-	Res. Seq. View	Top50BlastHits	BAA20761.1	Not given	Homo sapiens	PRI	16.4%	unmaskedSW	5-114	1836-1950	78
-	Res. Seq. View	Top50BlastHits	AAA26154.1	Not given	Homo sapiens	PRI	13.2%	unmaskedSW	5-112	10-137	79
-	Res. Seq. View	Top50BlastHits	AAA26154.1		Homo sapiens	PRI	13.8%	unmaskedSW	5-113	4-131	79
-	Res. Seq. View	Top50BlastHits	AAFD3046.1	candidate tumor suppressor protein DICE1	Escherichia coli	BCT	16.7%	unmaskedSW	5-39	250-335	78
-	Res. Seq. View	Top50BlastHits	AAC74854.1	orf, hypothetical protein	Homo sapiens	PRI	16.2%	unmaskedSW	63-156	2-90	80
-	Res. Seq. View	Top50BlastHits	AAE60942.1	breast cancer suppressor candidate 1	Homo sapiens	PRI	18.9%	unmaskedSW	3-112	111-251	72
-	Res. Seq. View	Top50BlastHits	AAE67537.1	glycoprotein IIIa	Homo sapiens	PRI	19.8%	unmaskedSW	3-112	138-277	74
-	Res. Seq. View	Top50BlastHits	AAE52589.1	Not given	Homo sapiens	PRI	19.6%	unmaskedSW	3-112	138-277	74
-	Res. Seq. View	Top50BlastHits	AAE71380.1	platelet membrane glycoprotein IIIa beta subunit	Homo sapiens	PRI	19.6%	unmaskedSW	3-112	138-277	74
-	Res. Seq. View	Top50BlastHits	AAA35927.1	Not given	Homo sapiens	PRI	19.6%	unmaskedSW	3-112	138-277	74

1) 509 hits identified by Genome Threader only:

FIG. 2B

4/40

Help

http://london-brIDGE.lnpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl

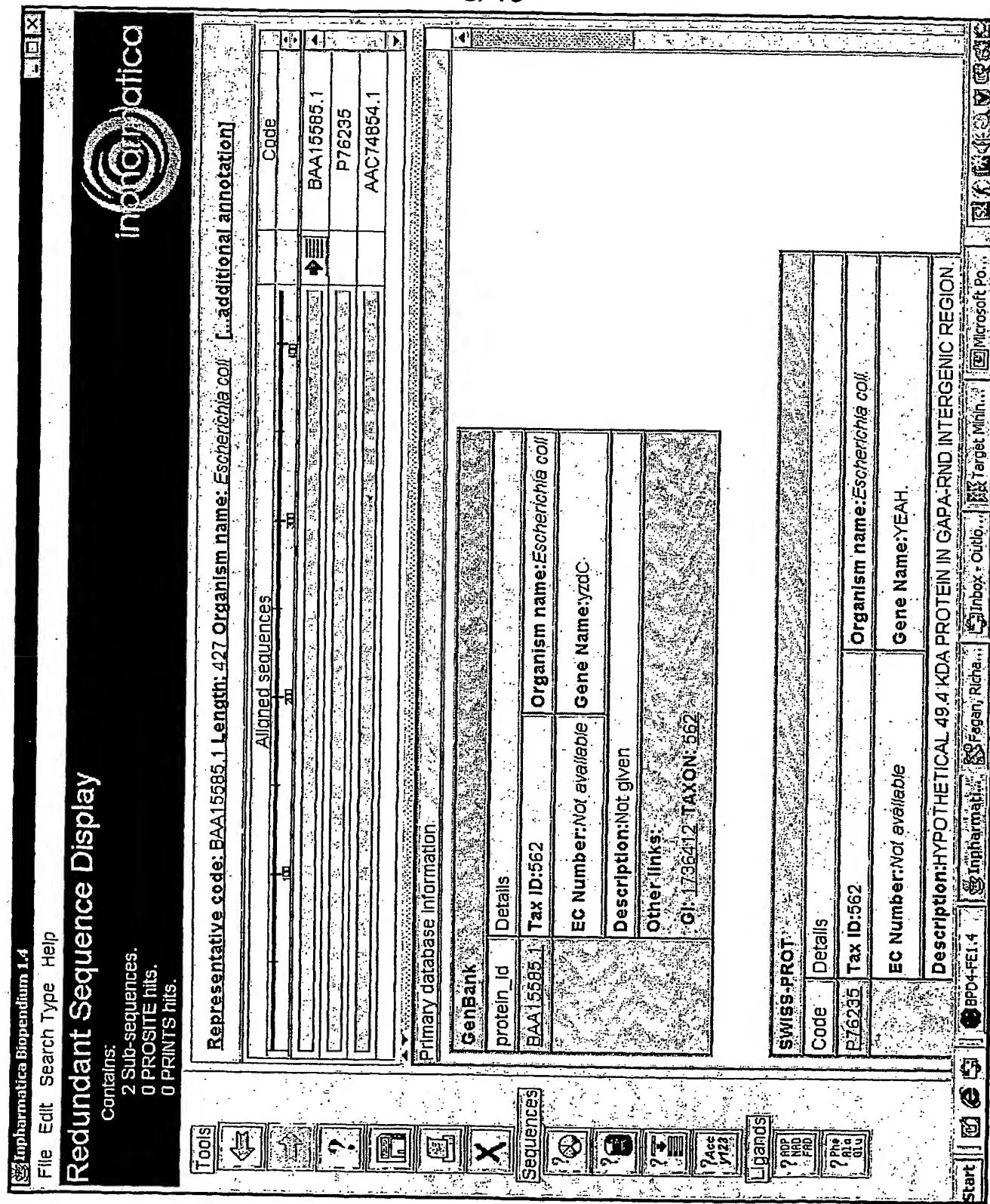
630 out of these 645 PSI-BLAST matches were identified using 'positive iterations':

15 out of these 645 PSI-BLAST matches were identified using 'negative iterations'.

A2) Genome Threader Matches:

FIG. 2C

5/40



୩୮

6/40

FIG. 4

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location: <http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi>

The Sanger Centre **Pfam**
Protein families database of alignments and HMMs

Home Keyword search Protein search DNA search Browse Pfam Taxonomy search Help

Results for gi|1788084|gb|AAC74854.1|

There were no matches to Pfam-A (including borderline matches) for gi|1788084|gb|AAC74854.1|

Matches to Pfam-B

Domain	Start	End	Eval	Alignment
Pfam-B_39416	233	423	3.7e-103	Align

[427 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: [Hypertext linked to swisspfam](#)

Query gi|1788084|gb|AAC74854.1|/233-423 matching [Pfam-B_39416](#)

```

YEAH_ECOLI 233 DLRYKNYEKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSR 282
DLRYKNYEKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSR
gi|1788084|gb|AAC74854.1| 233 DLRYKNYEKRPDPSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSR 282

YEAH_ECOLI 283 TYKNVEVYVYIRDMTQRAKEVDEMEFFYSQETGGTIUSSALXLMDEUVKERY 332
TYKNVEVYVYIRDMTQRAKEVDEMEFFYSQETGGTIUSSALXLMDEUVKERY
gi|1788084|gb|AAC74854.1| 283 TYKNVEVYVYIRDMTQRAKEVDEMEFFYSQETGGTIUSSALXLMDEUVKERY 332

YEAH_ECOLI 333 NPAQWNIYARQASDGDNWADDSPLCHEILAKKLLPVVRYYSYIEITRRAM 382
NPAQWNIYARQASDGDNWADDSPLCHEILAKKLLPVVRYYSYIEITRRAM
gi|1788084|gb|AAC74854.1| 333 NPAQWNIYARQASDGDNWADDSPLCHEILAKKLLPVVRYYSYIEITRRAM 382

YEAH_ECOLI 383 QTLWREYEMLQSTFDNRFAMQHIRDQDDIYVFRELHKQNR 423
QTLWREYEMLQSTFDNRFAMQHIRDQDDIYVFRELHKQNR
gi|1788084|gb|AAC74854.1| 383 QTLWREYEMLQSTFDNRFAMQHIRDQDDIYVFRELHKQNR 423

```

[Align to family](#)

If you think there is anything wrong with this script, please contact [Pfam](#)

FIG. 5

7/40

File Edit View Go Communicator Help

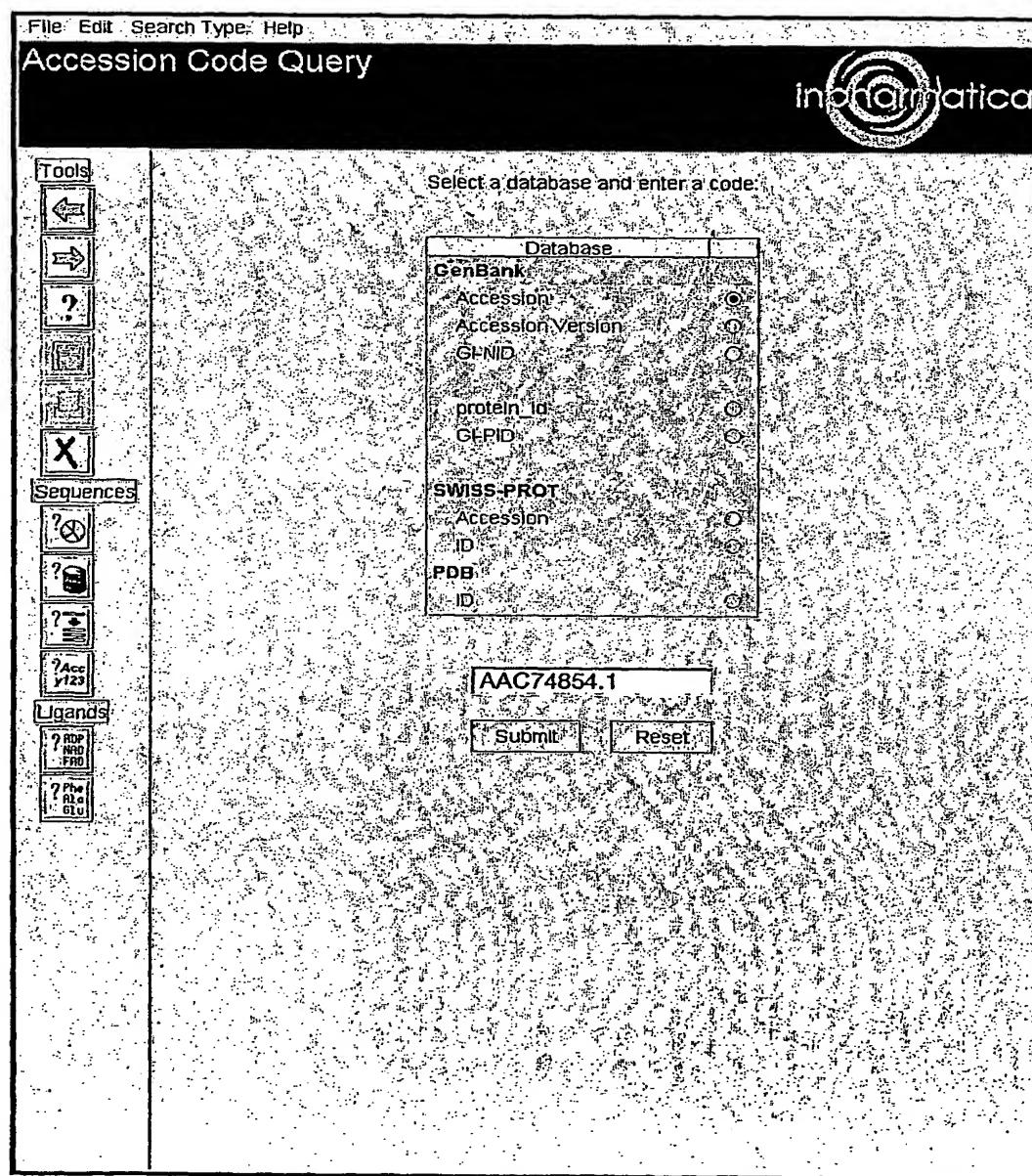
Bookmarks Location <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=...>

LOCUS	RAC74854	427 aa	BCT	01-DEC-2000
DEFINITION	orf, hypothetical protein [Escherichia coli K12].			
ACCESSION	RAC74854			
PID	gi1788084			
VERSION	RAC74854.1	gi:1788084		
DBSOURCE	locus	RE000273 accession <u>RE000273_1</u>		
KEYWORDS				
SOURCE	Escherichia coli K12.			
ORGANISM	Escherichia coli K12			
REFERENCE	1 (residues 1 to 427)			
AUTHORS	Blattner, F.R., Plunkett, G., III, Bloch, C.R., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.R., Rose, D.J., Mau, B. and Shao, Y.			
TITLE	The complete genome sequence of Escherichia coli K-12			
JOURNAL	Science 277 (5331), 1453-1474 (1997)			
MEDLINE	97426617			
PUBMED	9278503			
REFERENCE	2 (residues 1 to 427)			
AUTHORS	Blattner, F.R.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459			
REFERENCE	3 (residues 1 to 427)			
AUTHORS	Blattner, F.R.			
TITLE	Direct Submission			
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459			
REFERENCE	4 (residues 1 to 427)			
AUTHORS	Plunkett, G. III.			
TITLE	Direct Submission			
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA			
COMMENT	This sequence was determined by the <i>E. coli</i> Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from <i>E. coli</i> K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with <i>cg</i> Site Nos., unique ID nos. for the genes in the <i>E. coli</i> Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berklyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the <i>E. coli</i> Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The <i>E. coli</i> K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.			
FEATURES	Method: conceptual translation.			
source	Location/qualifiers			
	1..427			
	/organism="Escherichia coli K12"			
	/strain="K12"			
	/sub_strain="MG1655"			
	/db_xref="taxon:83333"			
Protein	1..427			
	/function="orf; Unknown"			
	/product="orf, hypothetical protein"			
CDS	1..427			
	/gene="peaf"			
	/coded_by="1788078:6385..7668"			
	/transl_table=11			
	/note="0427: This 427 aa ORF is 28 pct identical (43 gaps) to 327 residues of an approx. 312 aa protein YZDC_PRC5U SW: P45742"			
ORIGIN	1 mtwfidrrin gknksmvnrq rflrrykaqi kqsisceaink rsvtdvdsge svsiptedis 61 epmfhqgrgg lxhrvhpgmd hfvgmdrier pqqggggsgs gqqqasqdg e qddefvqfis 121 kdeyldilfe dlapnlkgn qqrqlteykt hragytangv panisvvrsl qnsllarrtam 181 taqkrrelha leenlaiish sepaqlleee rlrkeiaelr akiervpfd t f d l r y k n y e			

100%

8/40

FIG. 6A



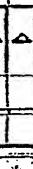
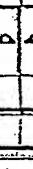
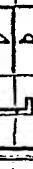
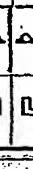
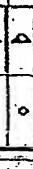
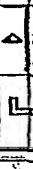
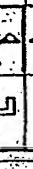
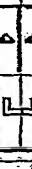
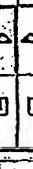
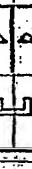
9/40

bioinformatica

File Edit Search Type Help

Aligned Sequence Display

Query details:
AAC74854.1: orf, hypothetical protein

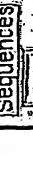
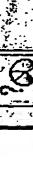
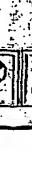
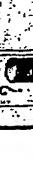
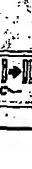
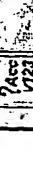
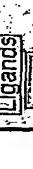
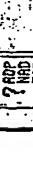
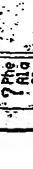
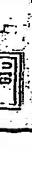
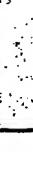
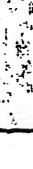
Topics                        

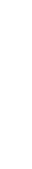
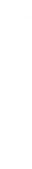
Page 1 of 2 Total hits: 36 Selected: 0

Total selected: 0 View alignment

Inpharmatica Genome Threader results:

Cluster Detail...	Code	Title	Organism	%ID	Query len.	Target len.	Aln. score	Method	Confidence
	1LFAA	CD11A I-DOMAIN WITH BOUND MN++	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(89%)
►	1Z00A	CD11A I-DOMAIN WITH BOUND MAGNESIUM	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(89%)
►	1Z00B	CD11A I-DOMAIN WITH BOUND MAGNESIUM	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(89%)
►	1Z0P8	CD11A I-DOMAIN WITH BOUND MAGNESIUM	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(89%)
►	1Z0PA	CD11A I-DOMAIN WITH BOUND MAGNESIUM	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal(89%)
►	1BH01	MAC-1 I DOMAIN MAGNESIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 335	4 - 128	76	Local	Low(84%)
►	1IDN1	MAC-1 I DOMAIN METALFREE	<i>H.sapiens</i>	15.1	250 - 335	4 - 128	76	Local	Low(84%)
►	1BH02	MAC-1 I DOMAIN CADMIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 335	4 - 128	76	Local	Low(84%)
►	1IDN2	MAC-1 I DOMAIN METALFREE	<i>H.sapiens</i>	15.1	250 - 335	4 - 128	76	Local	Low(84%)
►	1ATZB	HUMAN VON WILLEBRAND FACTOR A3 DOMAIN	<i>H.sapiens</i>	15.1	244 - 332	1 - 126	70	Local	Low(94%)
►	1ATZA	HUMAN VON WILLEBRAND FACTOR A3 DOMAIN	<i>H.sapiens</i>	15.3	250 - 335	5 - 101	72	Local	Low(94%)
►	1DGQA	NMR SOLUTION STRUCTURE OF THE INSERTED DOMAIN FROM INTEGRIN CR3 MIG2+ BO...	<i>H.sapiens</i>	16.3	250 - 331	9 - 89	75	Local	Low(93%)
►	1IDO	I-DOMAIN FROM INTEGRIN CR3 MN2+ BO...	<i>H.sapiens</i>	15.1	250 - 335	4 - 128	75	Local	Low(82%)
►	1JLM	I-DOMAIN FROM INTEGRIN CR3 MN2+ BO...	<i>H.sapiens</i>	15.1	250 - 335	4 - 128	75	Local	Low(89%)
►	1Z0N	CD11A I-DOMAIN WITHOUT BOUND CAT...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	72	Local	Low(83%)
►	1LFAB	CD11A I-DOMAIN WITH BOUND MN++	<i>H.sapiens</i>	18.0	250 - 333	5 - 97	66	Local	Low(80%)
►	1OAKA	CRYSTAL STRUCTURE OF THE VON WILLEBRAND FACTOR A3 DOMAIN OF VON WILLEBRAND FACTOR A1	<i>H.sapiens</i>	20.3	251 - 301	1 - 58	72	Local	Low(72%)
►	1AUG	A1 DOMAIN OF VON WILLEBRAND FACTOR A3 DOMAIN OF VON WILLEBRAND FACTOR A1	<i>H.sapiens</i>	21.2	250 - 301	18 - 68	69	Local	Low(66%)
►	1AO8A	A3 DOMAIN OF VON WILLEBRAND FACTOR A3 DOMAIN OF VON WILLEBRAND FACTOR A1	<i>H.sapiens</i>	15.2	245 - 331	2 - 99	63	Local	Low(53%)
►	1AO8B	A3 DOMAIN OF VON WILLEBRAND FACTOR A3 DOMAIN OF VON WILLEBRAND FACTOR A1	<i>H.sapiens</i>	15.2	245 - 331	2 - 99	63	Local	Low(53%)
►	1AOX8	I DOMAIN FROM INTEGRIN ALPHAI2-BETA1	<i>H.sapiens</i>	18.1	242 - 335	1 - 105	53	Local	Low(21%)
►	1RYPD	CRYSTAL STRUCTURE OF THE 20S PROTEASOME	<i>S.cerevisiae</i>	18.1	246 - 305	166 - 226	58	Local	Low(21%)
►	1RYPR	CRYSTAL STRUCTURE OF THE 20S PROTEASOME	<i>S.cerevisiae</i>	18.1	246 - 305	166 - 226	58	Local	Low(21%)
►	1CK4B	CRYSTAL STRUCTURE OF RAT ATB1 INTEGRIN	<i>Rattus norvegicus</i>	10.5	250 - 331	5 - 86	55	Local	Low(20%)

Sequences                   

Ligands                   

GenBank SWISS-PROT PDB

FIG. 6B

SUBSTITUTE SHEET (RULE 26)

10/40

Aligned Sequence Display

Query details:
AAC74854.1: orf, hypothetical protein

Total selected: 0

View alignment

Reverse Maximised PSI-BLAST results.

Total hits: 2 Selected: 0

Page 1 of 1

Cluster Detail Code

Cluster	Detail	Code	Title	Organism	%ID	Query len.	Target len.	Blter.	E value
0	0	CAB12725.1	Not given	<i>B. subtilis</i>	23.0	14 - 423	21 - 385	1	2.00E-29
0	0	AAB8880.01	conserved hypothetical protein	<i>M. jannaschii</i>	15.0	234 - 371	259 - 401	-8	4.00E-09

Tools

GenBank SWISS-PROT PDB

FIG. 6C Aligned Sequence Display

FIG. 7

AllEye output (January 4, 2002 3:07 PM)

11/40

BAA15585.1 MTWFIDRRRLNGKNKSMVNRAFLARRYKAQIKOSISEAINKRSVTDVOSGESVSIPTEDISEPMFHQGRGGGLRHAVHPPGN
 1LFAA

BAA15585.1 DHFVQNDRIERPQGGGGSGGGQGQASQDGEQGDEFVfqISKDDEYLDLFLFEDLALPNLKQNQQRQLTEYKTHRAGYTAN
 1LFAA

BAA15585.1 GVPANISVVRSLQNSLARRTAMTAGKRELHAAEENLAIISSEPAQLLEEEERLKEIAELRAKIERVPPFIDTFDLRYK
 1LFAA

BAA15585.1 NYEKRPDPSSQAVMFCLMDVGSMDQSTKDMAKRFYILLYLFLSRTYKNVEV...YIRHHTQAKEVDEEEF
 1LFAA

BAA15585.1 QNVDLVF.LFDGSMMSLQPDFEQKILDFMKVMKLISNTSYQFAAVQFSTSYKTEFDFFSDYVVKRKDPDALLK
 1LFAA

BAA15585.1 FYSQETGGTIVSSALKLMDEVVKERYNPAQWNLYAAQASDGDNWADDSPLCHEILAKKLLPVVRYYSYIEITRAA.HQT
 1LFAA

BAA15585.1 HVKHMLLTNTFGAINYVATEVFREELGARPDATKVLLITDGEATD...SGNIDAAKDIIIRY..IIGIGKHFQTK
 1LFAA

BAA15585.1 LWREYEHLQSTFDNFAMQHIRDQDDIYPVFRELFHQNATAKG
 1LFAA

SQETLHKFASKPASEFVKILDFFEKLKDLFTE.LQKKIYVIE

SUBSTITUTE SHEET (RULE 26)

FIG. 8B

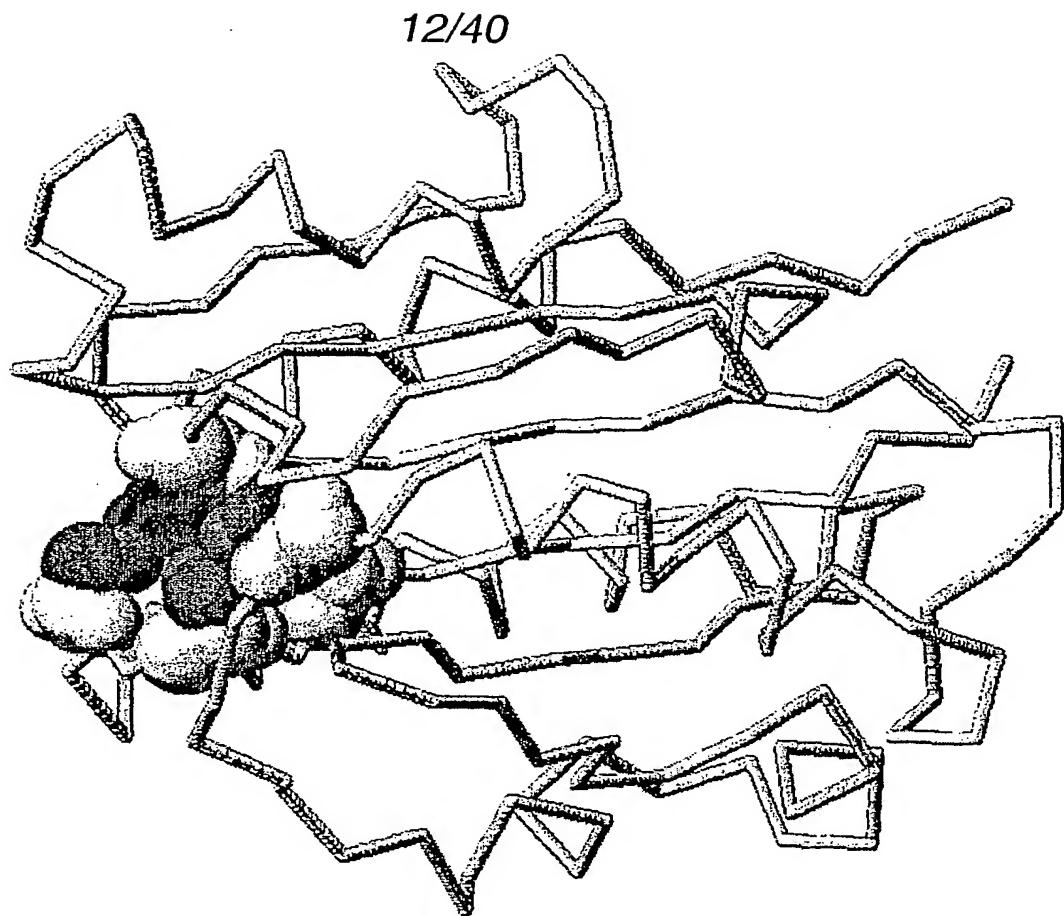
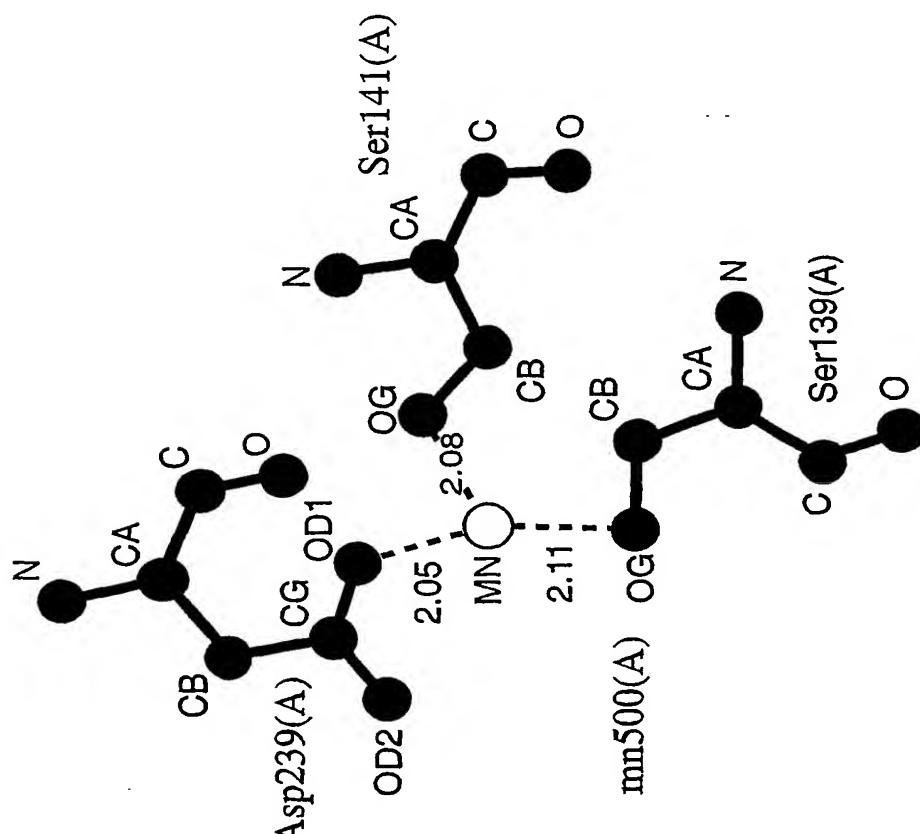


FIG. 8A

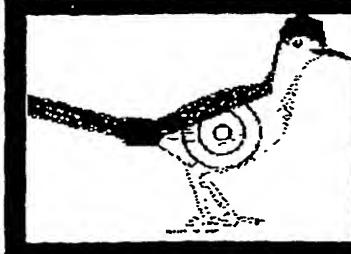


11fa: MN500 Chain [A]

FIG. 9

13/40

File Edit View Go Communicator Help
Bookmarks Location: http://victoria.inpharmatica.co.uk/~volker/BPD3target.html

 Target Mining Interface 

Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

14/40

Combined Genome Threader and PSI-BLAST output PSI-BLAST values are shown in parentheses!

2) 82 additional hits identified by both, Genome Threader and PSI-BLAST:

Accession	EGP400:	WWW-Ent:	Title:	Organism	Env.	%ID (GT,PSI)	Query len. (GT,PSI)	Target len. (GT,PSI)	Aln. score	Conf. (GT)	1st iter. (PSI)	Best iter. (PSI)	Best E-value (PSI)
EE6192			INTEGRIN ALPHAI-1 (LAMININ AND COLLAGEN RECEPTOR) (VLA-1) (CD9A).	Homo sapiens (Human).	PR1	51.8%, 51%	2-200, 2-200	139-337, 139-337	487	100%	1	2	2E-73
EE7301			PLATELET MEMBRANE GLYCOPROTEIN IA PRECURSOR (GP1A) (COLLAGEN RECEPTOR) (INTEGRIN ALPHAI-2) (VLA-2/ALPHA CHAIN) (CD9B).	Homo sapiens (Human).	PR1	55.3%, 100%	2-200, 2-200	169-387, 169-387	471	100%	1	1	1E-111
AA012581			Integrin alpha-11 subunit precursor	Homo sapiens	PR1	45.2%, 45%	2-200, 2-200	159-355, 159-355	469	100%	1	2	4E-68
AA012581			Integrin alpha 11 subunit precursor	Homo sapiens	PR1	45.2%, 45%	2-200, 2-200	159-355, 159-355	468	100%	1	2	4E-68
AA0519182			Integrin alpha 11 subunit precursor	Homo sapiens (Human).	PR1	28.1%, 31%	2-188, 7-189	133-318, 2323-2495	437	100%	1	2	2E-42
Q89215			COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PR1	27.1%, 27%	2-200, 2-200	144-338, 144-338	435	100%	1	2	1E-60
AA0519182			COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PR1	27.1%, 27%	2-200, 2-200	144-338, 144-338	435	100%	1	2	1E-60
AAA325241			Not given	Homo sapiens	PR1	27.1%, 27%	2-200, 2-200	144-338, 144-338	435	100%	1	2	1E-60
AAA325241			Not given	Homo sapiens	PR1	27.1%, 27%	2-200, 2-200	144-338, 144-338	435	100%	1	2	1E-60
CA2712221			cartilage	Homo sapiens	PR1	26.5%, 26%	158-201	233-412, 1581-1626	424	100%	2	3	5E-6
EE21941			CARTILAGE MATRIX PROTEIN PRECURSOR (MATRILIN-1).	Homo sapiens (Human).	PR1	26.0%, 26%	3-188, 3-188	271-485, 37-224	424	100%	1	2	3E-46
AA015061			Cartilage	Homo sapiens	PR1	27.6%, 30%	2-185, 2-185	133-318, 133-318	422	100%	1	2	3E-50
AA015061			type XII collagen	Homo sapiens	PR1	46.5%, 45%	2-188, 2-199	162-359, 162-359	419	100%	1	2	5E-63
AA015061			Integrin subunit alpha 10 precursor	Homo sapiens	PR1	26.2%, 26%	1-189, 2-200	335-529, 25-223	411	100%	1	2	1E-47
CA015591			matrin-1	Homo sapiens	PR1	26.2%, 26%	1-189, 2-200	335-529, 25-223	411	100%	1	2	1E-47

FIG. 10A

SUBSTITUTE SHEET (RULE 26)

15/40

FIG. 10B

608 out of these 632 PSI-BLAST matches were identified using 'positive iterations':

FIG. 10C

24 out of these 632 PSI-BLAST matches were identified using 'negative iterations'.

17/40

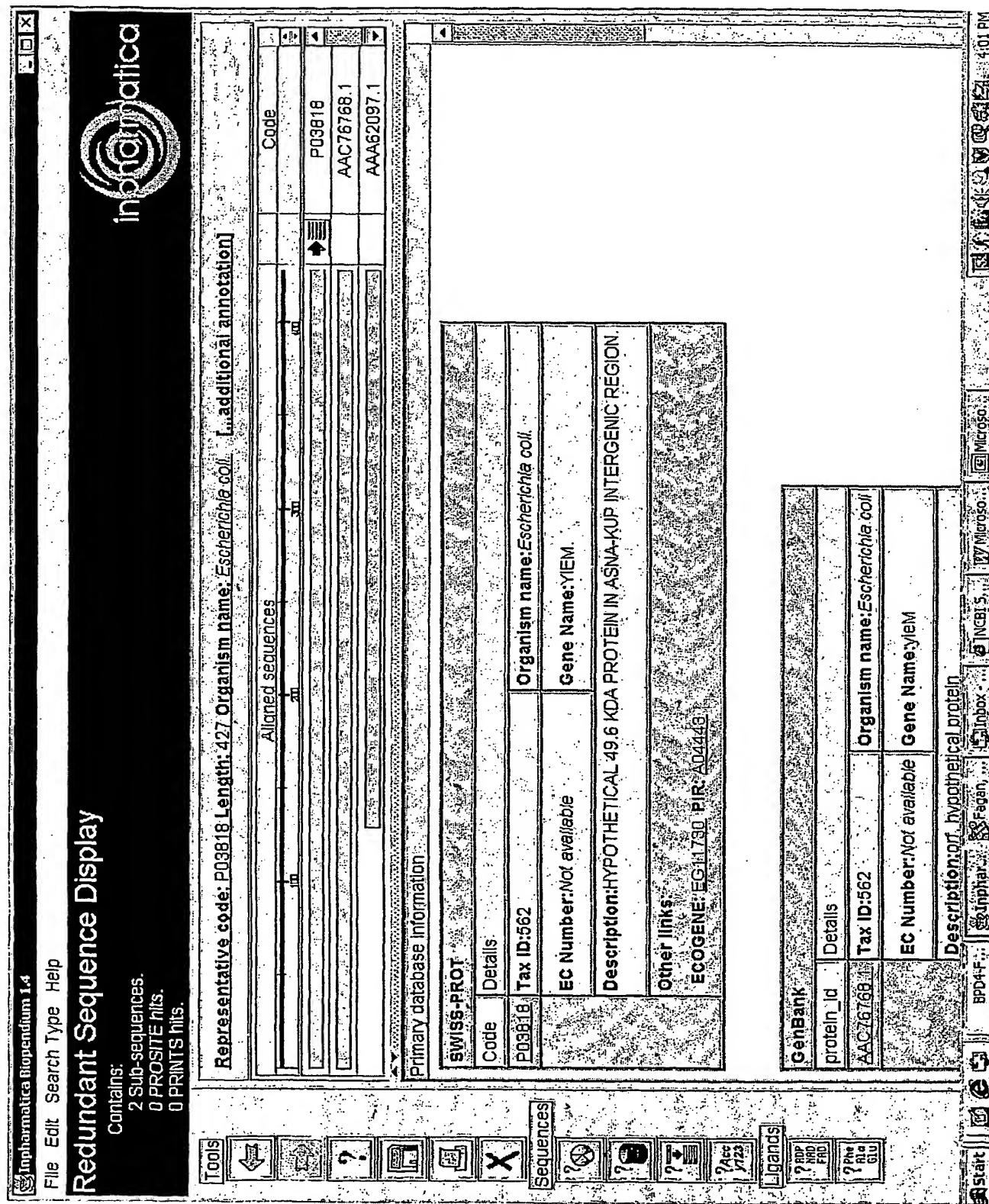


FIG. 1

18/40

FIG. 12

File Edit View Go Communicator Help
 Bookmarks Location: <http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi>

The Sanger Centre **Pfam**
 Protein families database of alignments and HMMs
[Home](#) | [Keyword search](#) | [Protein search](#) | [DNA search](#) | [Browse Pfam](#) | [Taxonomy search](#) | [Help](#)

Results for gi|2367274|gb|AAC76768.1|

There were no matches to Pfam-A (including borderline matches) for gi|2367274|gb|AAC76768.1|

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B 15204	204	408	2.4e-108	Align

[427 residues]

.....

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: Hypertext linked to swisspfam

Query gi|2367274|gb|AAC76768.1|/204-408 matching Pfam-B 15204

```

YIEV_ECOLI 204 DILRLPPPELATLGITELEYEFYRRLVEKQLLTYRLNGESWREKIERPV 253
          DILRLPPPELATLGITELEYEFYRRLVEKQLLTYRLNGESWREKIERPV
gi|2367274|gb|AAC76768.1| 204 DILRLPPPELATLGITELEYEFYRRLVEKQLLTYRLNGESWREKIERPV 253

YIEV_ECOLI 254 VVKDYDEQPRGPFIVCVDTSGSMGFPNEQCAKAPCLRLMRILALAEHRRCT 303
          VVKDYDEQPRGPFIVCVDTSGSMGFPNEQCAKAPCLRLMRILALAEHRRCT
gi|2367274|gb|AAC76768.1| 254 VVKDYDEQPRGPFIVCVDTSGSMGFPNEQCAKAPCLRLMRILALAEHRRCT 303

YIEV_ECOLI 304 IMLPSTEIVRYELSGPQGIEQAIRFLSQQFRGCTDLASCYRAIMERLQSR 353
          IMLPSTEIVRYELSGPQGIEQAIRFLSQQFRGCTDLASCYRAIMERLQSR
gi|2367274|gb|AAC76768.1| 304 IMLPSTEIVRYELSGPQGIEQAIRFLSQQFRGCTDLASCYRAIMERLQSR 353

YIEV_ECOLI 354 EWPDRDAVVISDFIQLPDDVTSKVELQRVHQMRPNNRVRMSRMGKPGI 403
          EWPDRDAVVISDFIQLPDDVTSKVELQRVHQMRPNNRVRMSRMGKPGI
gi|2367274|gb|AAC76768.1| 354 EWPDRDAVVISDFIQLPDDVTSKVELQRVHQMRPNNRVRMSRMGKPGI 403

YIEV_ECOLI 404 MRIFD 408
          MRIFD
gi|2367274|gb|AAC76768.1| 404 MRIFD 408
  
```

.....

If you think there is anything wrong with this script, please contact Pfam

FIG. 13

19/40

File Edit View Go Communicator Help

Bookmarks Location: http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=protein

LOCUS ARCT76768 427 aa BCT 01-DEC-2000

DEFINITION orf, hypothetical protein [Escherichia coli K12].

ACCESSION ARCT76768

PID 92367274

VERSION ARCT76768.1 cr:2367274

DBSOURCE locus AE000451 accession AE000451.1

KEYWORDS

SOURCE Escherichia coli K12.

ORGANISM Escherichia coli K12

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (residues 1 to 427)

AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.R., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, R.W., Kirkpatrick, H.R., Goeden, M.R., Rose, D.J., Mau, B. and Shao, Y.

TITLE The complete genome sequence of Escherichia coli K-12

JOURNAL science 277 (5331); 1453-1474 (1997)

MEDLINE 97426617

PUBMED 9278503

REFERENCE 2 (residues 1 to 427)

AUTHORS Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 3 (residues 1 to 427)

AUTHORS Blattner, F.R.

TITLE Direct Submission

JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 4 (residues 1 to 427)

AUTHORS Blattner, G. III.

TITLE Direct Submission

JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

COMMENT This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG000301 and HG03428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using geneWarp software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli genetic Stock center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source Method: conceptual translation.

Location/Qualifiers 1..427

/organism="Escherichia coli K12"

/strain="K12"

/sub_strain="MG1655"

/db_xref="taxon:83333"

Protein 1..427

/function="orf; Unknown"

/product="orf, hypothetical protein"

cds 1..427

/gene="yieM"

/coded_by="complement(2367272:5249..6532)"

/transl_table=11

/note="f427; sequence change joins ORFs yieM and yiem from earlier version"

ORIGIN

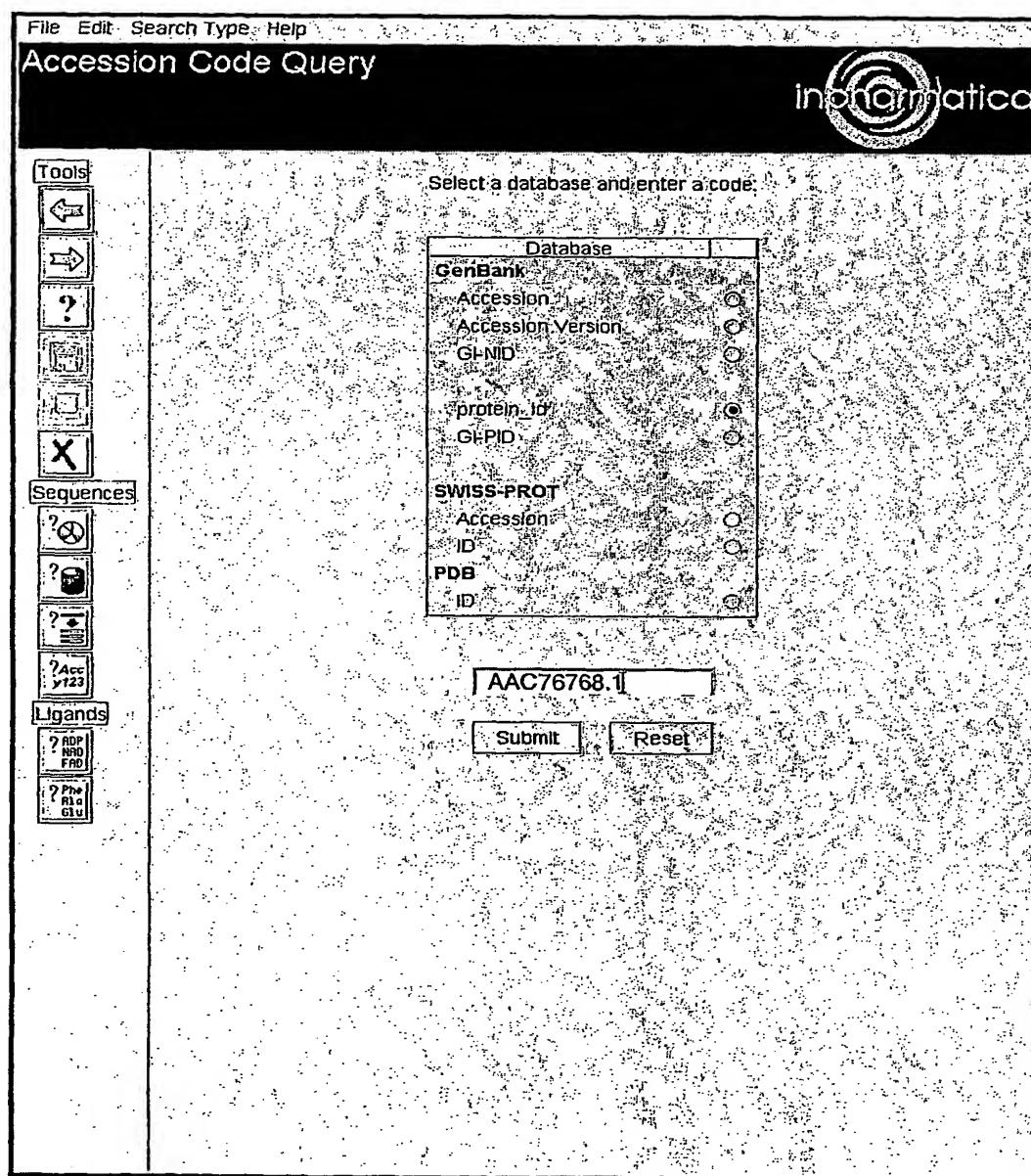
1 mrsrlkdarv ppelteewmc pqqsqqlstp qfivqlpqil dlhrlnspw aegargqlvda
 61 nstitsalht lflqqrwlrl iqvqattinq lleereqell sevgemtis gglepiladm
 121 ntaagrlwda saggqlkrqdy qllvkpgefl neppelkrla eqlgrsreak siphndmge
 181 tftrtmvrepq tpeqgvdgllq qsdldilrlb pealgtita lonefuxqiv sklltq-1h



100%

20/40

FIG. 14A



21/40

FIG. 14B Aligned Sequence Display

Query details: AAC76768.1: orf, hypothetical protein

Query details:

File Edit Search Type Help

Aligned Sequence Display

Query details: AAC76768.1: orf, hypothetical protein

Total selected: 0.

Page 1 of 3 Total hits: 88. Selected: 0.

Inpharmatica Genome Threader results:

View alignment

Organism: *Ranunculus*

Method: Local

Confidence: Certain(100%)

Cluster Detail.. Code Title %ID Query.qn. Target.qn. Alt.score Method Confidence

1 1CK4B CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN 12.6 267 - 398 6 - 144 158 Local Certain(100%)

2 1IDO I-DOCK FROM INTEGRIN C13, MG2+ BO... *H.sapiens* 19.8 267 - 384 5 - 130 157 Local Certain(100%)

3 1JLM I-DOCK FROM INTEGRIN C13, MN2+ BO... *H.sapiens* 19.8 267 - 384 5 - 130 156 Local Certain(100%)

4 1CK4A CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN 12.6 267 - 398 6 - 144 149 Local Certain(100%)

5 1BHQ1 MAC-1 DOMAIN MAGNESIUM COMPLEX 19.8 267 - 384 6 - 130 146 Local Certain(100%)

6 1BHQ2 MAC-1 DOMAIN MAGNESIUM COMPLEX 19.8 267 - 384 5 - 130 146 Local Certain(100%)

7 1BHQ1 MAC-1 DOMAIN CADMIUM COMPLEX 19.8 267 - 384 5 - 130 146 Local Certain(100%)

8 1IDN2 MAC-1 DOMAIN IN METAL FREE... 19.8 267 - 384 5 - 130 146 Local Certain(100%)

9 1IDN1 MAC-1 DOMAIN METAL FREE 19.8 267 - 384 5 - 130 146 Local Certain(100%)

10 1BHQ2 MAC-1 DOMAIN MAGNESIUM COMPLEX 19.8 267 - 384 5 - 130 146 Local Certain(100%)

11 1LFA8 CD11A I-DOCK WITH BOUND MN++ *H.sapiens* 15.9 264 - 381 1 - 124 149 Local Certain(100%)

12 1ZON CD11A I-DOCK WITHOUT BOUND CAT... *H.sapiens* 15.9 264 - 381 1 - 124 143 Local Certain(100%)

13 1LFAA CD11A I-DOCK WITH BOUND MN++ *H.sapiens* 14.6 264 - 354 1 - 86 141 Local Certain(100%)

14 1ZQ0A CD11A I-DOCK WITH BOUND MAGNESI... *H.sapiens* 14.6 264 - 354 1 - 86 141 Local Certain(100%)

15 1ZQ0B CD11A I-DOCK WITH BOUND MAGNESI... *H.sapiens* 14.6 264 - 354 1 - 86 141 Local Certain(100%)

16 1ZQ0A CD11A I-DOCK WITH BOUND MAGNESI... *H.sapiens* 14.6 264 - 354 1 - 86 141 Local Certain(100%)

17 1ZQ0B CD11A I-DOCK WITH BOUND MAGNESI... *H.sapiens* 14.6 264 - 354 1 - 86 141 Local Certain(100%)

18 1DGQ4 NMR SOLUTION STRUCTURE OF THE INTEGRIN ALPH1 BETA1 SUBUNIT 14.3 267 - 354 10 - 100 139 Local Certain(100%)

19 1ATZA HUMAN VON WILLEBRAND FACTOR A3 DO... *H.sapiens* 11.3 267 - 407 6 - 152 135 Local Certain(100%)

20 1OAKA CRYSTAL STRUCTURE OF THE VON WILLEBRAND FACTOR A3 DOCKING DOMAIN 11.3 267 - 398 10 - 150 135 Local Certain(100%)

21 1ACKA I-DOCK DOMAIN FROM INTEGRIN ALPH2-BETA1 13.2 267 - 384 9 - 134 107 Local Certain(100%)

22 1QGCEB I-DOCK DOMAIN FROM INTEGRIN ALPH2-BETA1 11.8 267 - 398 5 - 143 106 Local Certain(100%)

23 1QGCA A1 DOMAIN FROM INTEGRIN ALPH2-BETA1 N/A 11.8 267 - 398 7 - 145 102 Local Certain(100%)

24 1AOXB I-DOCK DOMAIN FROM INTEGRIN ALPH2-BETA1 *H.sapiens* 13.3 267 - 398 6 - 149 98 Local Certain(100%)

25 1AOXA A2 DOMAIN OF VON WILLEBRAND FACTOR 10.7 267 - 398 8 - 145 98 Local Certain(100%)

26 1ACGIB A3 DOMAIN OF VON WILLEBRAND FACTOR *H.sapiens* 10.7 267 - 398 8 - 145 98 Local Certain(100%)

27 1AUG A1 DOMAIN OF VON WILLEBRAND FACTOR *H.sapiens* 11.8 250 - 398 4 - 159 97 Local Certain(100%)

28 1ATZB HUMAN VON WILLEBRAND FACTOR A3 DOCKING DOMAIN 10.7 267 - 398 8 - 146 96 Local Certain(100%)

29 1AGLB NITROGEN REGULATORY BACTERIAL PRO... *E.coli* 18.6 306 - 349 13 - 57 80 Local Low(42%)

30 1BRDS COMPLEX OF HUMAN-N-RAS WITH HUMAN *H.sapiens* 14.4 147 - 237 100 - 202 138 Local Low(34%)

Sequences

Ligands

Tools

SWISS-PROT

GenBank

SUBSTITUTE SHEET (RULE 26)

Aligned Sequence Display

Query details: AAC76768.1: orf, hypothetical protein

Total selected: 0 View alignment

Page 1 Total hits: 508 Selected: 0 Selected all Deseret all

Reverse Maximised PSI-BLAST results:

Cluster Detail	Code	Title	Organism	%ID	Query len	Target len	Biter.	E value
o	BAA80016.1	452aa long hypothetical protein	<i>Aenomix</i>	28.0	156 - 407	202 - 491	2	2.00E-60
o	Q80385	HYPOTHETICAL PROTEIN MJD077	<i>M. marmosellum</i>	22.0	146 - 406	122 - 361	3	2.00E-71
o	AAB98810.1	conserved hypothetical protein	<i>M. marmosellum</i>	23.0	147 - 410	170 - 420	3	2.00E-65
o	AACT9510.1	putative retroelement polyprotein	<i>A. thaliana</i>	16.0	168 - 404	144 - 403	3	2.00E-50
o	Q55874	HYPOTHETICAL 45.8 KDA PROTEIN SLL0103	<i>S.spp.</i>	21.0	241 - 407	17 - 188	3	4.00E-38
o	AABB8957.1	Magnesium chelatase subunit Chl	<i>Methyloautotrophicum</i>	25.0	151 - 370	324 - 535	3	5.00E-22
o	CAA30478.1	Not given	<i>M. musculus</i>	20.0	247 - 416	132 - 308	4	2.00E-44
o	AAA59844.1	Not given	<i>H. sapiens</i>	16.0	223 - 412	107 - 305	4	3.00E-42
o	AAB24821.1	leukocyte integrin alpha chain	<i>H. sapiens</i>	16.0	223 - 412	107 - 305	4	3.00E-42
o	1JEM	1- DOMAIN FROM INTEGRIN CR3, MN2+ BOUND	<i>H. sapiens</i>	16.0	263 - 412	1 - 168	4	2.00E-41
o	1ID0	1-DOMAIN FROM INTEGRIN CR3, MG2+ BOUND	<i>H. sapiens</i>	16.0	263 - 412	1 - 168	4	2.00E-41
o	1BHQ1	MAC-1 DOMAIN MAGNESIUM COMPLEX	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	1IDN2	MAC-1 DOMAIN METAL FREE	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	1IDN1	MAC-2 DOMAIN MAGNESIUM COMPLEX	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	1BHQ2	MAC-11 DOMAIN MAGNESIUM COMPLEX	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	1BHQ2	MAC-11 DOMAIN CADMIUM COMPLEX	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	1BHQ1	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. sapiens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	Q22457	MAGNESIUM-CHELATE-SE SUBUNIT CHL PRECURSO	<i>H. sapiens</i>	18.0	162 - 417	488 - 701	4	7.00E-41
o	CAB62268.1	Not given	unidentified	18.0	162 - 417	473 - 705	4	2.00E-40
o	AAB16869.1	CD1b	<i>S. scrofa</i>	17.0	254 - 417	6 - 177	4	8.00E-40
o	CAB50178.1	Mg- protoporphyrin X	<i>H. sapiens</i>	17.0	162 - 410	90 - 311	4	2.00E-39
o	AAD52031.1	Magnesium-chelatase subunit CHL precursor	<i>A. thaliana</i>	17.0	162 - 423	441 - 665	4	5.00E-38
o	BAA16787.1	Mg chelatase subunit	<i>S.spp.</i>	18.0	189 - 367	415 - 588	4	6.00E-38
o	AAB88828.1	conserved protein	<i>M. marmosellum</i>	13.0	154 - 383	50 - 303	4	7.00E-37
o	CAA77537.1	S61 aa (60 kD) Mg chelatase subunit	<i>R. capsulatus</i>	21.0	168 - 367	277 - 484	4	1.00E-36
o	AAF22885.1	27G7.20	<i>A. thaliana</i>	17.0	162 - 423	481 - 747	4	2.00E-36
o	AAF21241.1	alpha D integrin	<i>R. norvegicus</i>	14.0	239 - 412	126 - 307	4	2.00E-35
o	AAF23492.1	leukocyte adhesion glycoprotein p150/95 alpha integrin subunit	<i>M. musculus</i>	13.0	231 - 412	118 - 307	4	2.00E-35
o	AAAR0542.1	leukocyte alpha 1 chain	<i>H. sapiens</i>	13.0	239 - 410	124 - 303	4	3.00E-35

Tools:

Alignments:

Regions:

Proteins:

GenBank:

SWISS-PROT:

PDB:

FIG. 14C

FIG. 15

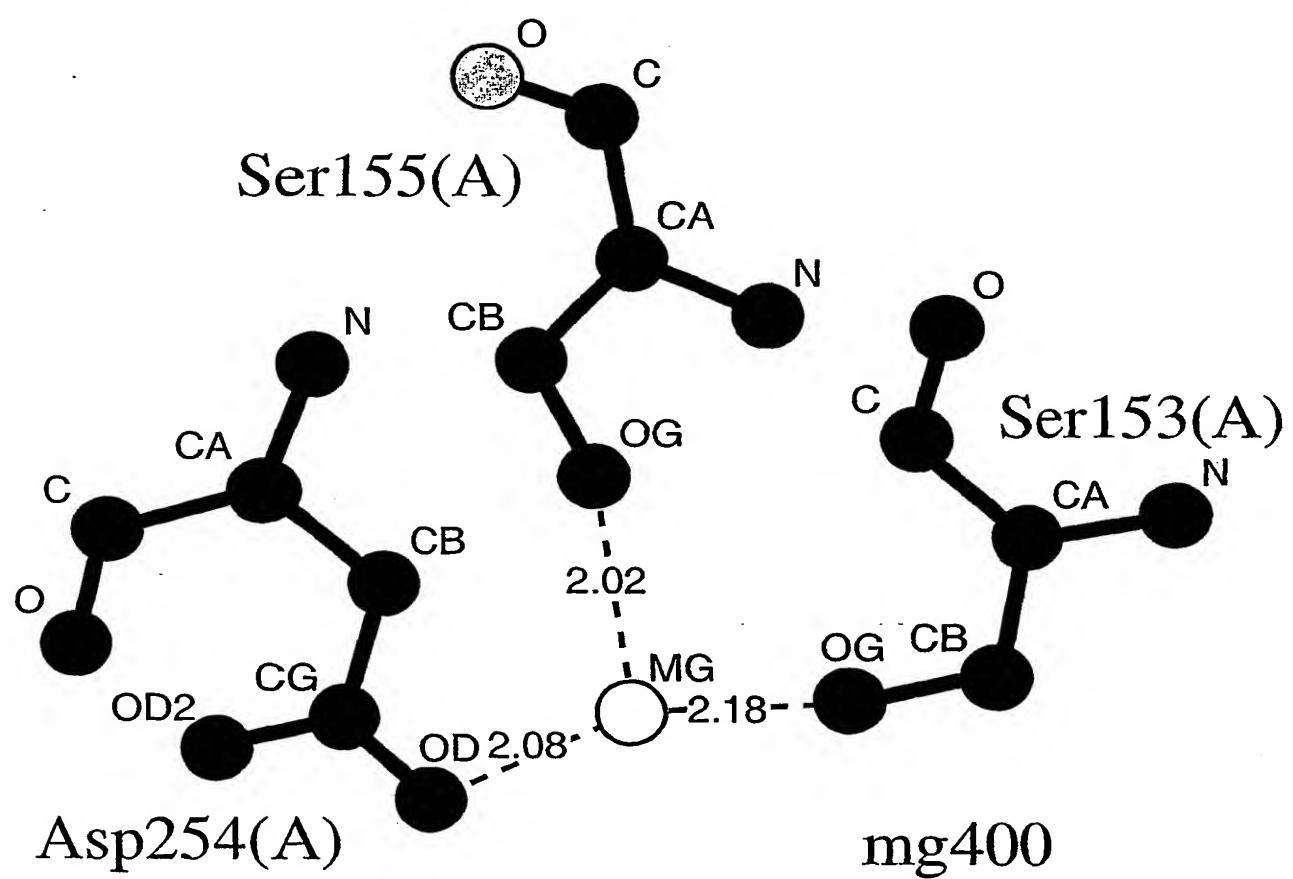
AIEye output (December 13, 2000 3:07 PM)

23/40

AAC76768.1	MASALKDARVPPPELTTEVINCYQQSSQLLSTPQFIVQLQPQILDLHLHPLNSPWAEQARQLYDANSTITSALHTLFLQRAWA	1AQX:R	10 20 30 40 50 60 70
AAC76768.1	LSLIVQATTLNQQQLLEEEEQQLSEYQERNTLSEQQLEPIIADNNNTAAGRALWDMSAGQLKAGSDYQLIVKYGEFLNEQP	1AQX:R	80 90 100 110 120 130 140 150
AAC76768.1	ELKRLAEGQLRSREAKSIPRNDQNETFRATMVRREPATVPEQVQDGLQQSDDLRLLPPELATLGITELYEFFYRPLVE	1AQX:R	160 170 180 190 200 210 220 230
AAC76768.1	KQLLTYRLHGESWREKYIERRPVVHKDYDEQPPGPFIVCVDTSGSMGGFNEQCAKAFCCLALMNR--ALAENRRCYIML	1AQX:R	240 250 260 270 280 290 300
AAC76768.1	SCPSLID--VYVYVCDENSIYPWDA--VKNFLEKFVQQLDIGPTKTQVGLI	1AQX:R	310 320 330 340 350 360 370 380
AAC76768.1	FSTEI-VAYEELSGPQQ--EQAIIRFLSQQQFRGGTDLASCFRAIMERLQS---REWFDAADVVISDFIAQRLPDD	1AQX:R	390 400 410 420 430 440 450 460
AAC76768.1	YANNPPVVFNLNTYKTKEEMIVATSAQSATSAQYGGDITNTFGAIQYAYAKYAYSAASGGRRRSGATKVMVVVTDGESEH-DGSMW	1AQX:R	470
AAC76768.1	VTTSKVKEELQAVHQRFHAYAMSKXHGKPGIMMRIFDHWRFDTGMWASRLLRAWA	1AQX:R	
AAC76768.1	LKAVIDQCNHDN-ILRFGIAVLGY-----LNRNA LDTKNLIKEIKAIASIPTERYFFNVSDERAKLEKAG	1AQX:R	
AAC76768.1	TLGEGQIFSIEGGT		

24/40

FIG. 16A



1aox: MG400

25/40

FIG. 16B

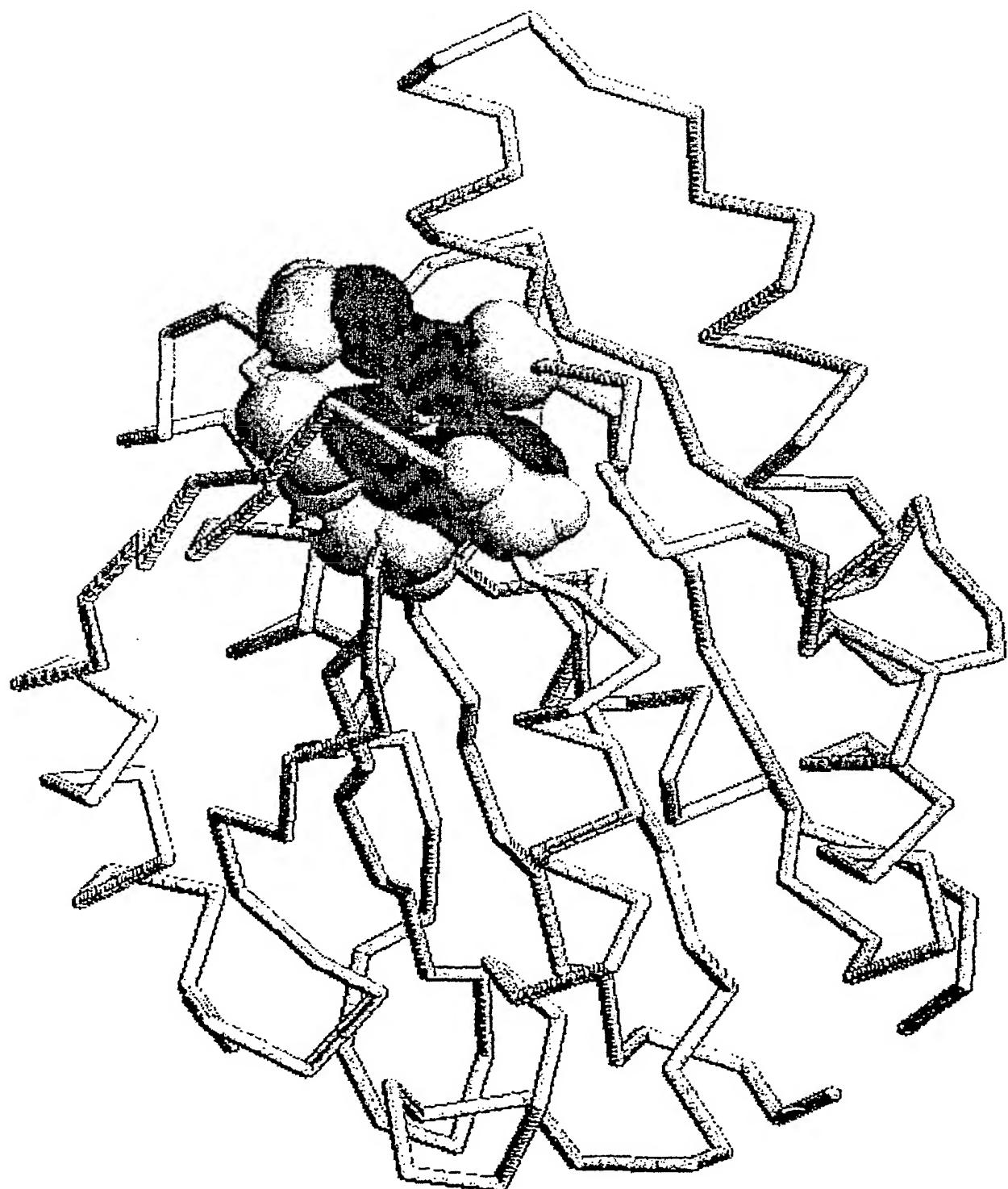
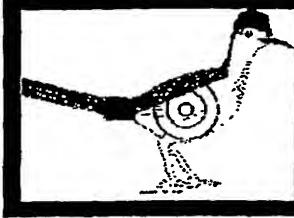


FIG. 17

26/40

File Edit View Go Communicator Help

Bookmarks Location: http://victoria.inpharmatica.co.uk/~volker/BPD3target.html

 Target Mining Interface



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

None

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:

Homo sapiens Rattus norvegicus (Rat) Mus musculus (Mouse) Danio rerio (Zebra fish)

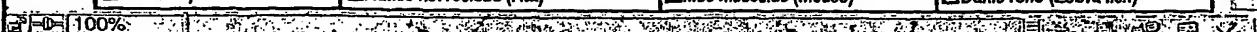
100% 

FIG. 18A

27/40

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location: http://London-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl

2) 81 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!

Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st Iter. (PSI)
↓	AAA59544.1 drill through Top50BlastHits	AAA59544.1	Not given	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	<u>100%</u> unmaskedGT	1
→	AAB24821.1 drill through Top50BlastHits	AAB24821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	468	<u>100%</u> unmaskedGT	1
↓	Q99715 drill through Top50BlastHits	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human)	PRI	28.9%, 26% unmaskedSW	2-186, 2-179	439-617, 2322-2494	456	<u>100%</u> unmaskedGT	1
↓	AAB38702.1 drill through Top50BlastHits	AAB38702.1	cartilage matrix protein	Homo sapiens	PRI	28.9%, 25% unmaskedSW	2-186, 2-186	274-452, 40-221	446	<u>100%</u> unmaskedGT	1
↓	AAC01506.1 drill through Top50BlastHits	AAC01506.1	type XII collagen	Homo sapiens	PRI	28.4%, 28% unmaskedSW	2-186, 2-186	137-318, 137-318	445	<u>100%</u> unmaskedGT	1
↓	CAA72402.1 drill through Top50BlastHits	CAA72402.1	collagen type XIV	Homo sapiens	PRI	28.7%, 30% unmaskedSW	2-186, 2-186	6-185, 6-185	442	<u>100%</u> unmaskedGT	1
↓	AAB38547.1 drill through Top50BlastHits	AAB38547.1	leukointegrin alpha d chain	Homo sapiens	PRI	61%, 60% unmaskedSW	1-187, 1-187	148-334, 148-334	439	<u>100%</u> unmaskedGT	1
↓	CAB71222.1 drill through Top50BlastHits	CAB71222.1	dJ238D15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	27.1%, 22% unmaskedSW	1-186, 2-186	293-472, 1430-1620	439	<u>100%</u> unmaskedGT	1
↓	CAA07569.1 drill through Top50BlastHits	CAA07569.1	matrin-4	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-187	344-525, 34-215	418	<u>100%</u> unmaskedGT	2
↓	CAB46380.1 drill through Top50BlastHits	CAB46380.1	dJ453C12.3 (matrin-4)	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-186	385-566, 34-214	418	<u>100%</u> unmaskedGT	2

28/40

FIG. 18B

Bookmarks		Location: http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPP3.pl		KIA0638 protein		Homo sapiens		PPI		unmaskedSW		73-185		390-503		655-700		unmaskedGT		reverse HI		
1	Red Seta View	SGH14009	BA031611			Homo sapiens		PPI		14.1%		6-88		20-105		76		89.6%				
1	Red Seta View	SGH14009	CA0521921	G7c protein		Homo sapiens		PPI		unmaskedSW												
1	Red Seta View	SGH14009	CA0521921	N-copine		Homo sapiens		PPI		26.8%		73-143		405-474		73		89.6%				
1	Red Seta View	SGH14009	CA0521921	INTEGRIN BETA-8 PRECURSOR		Homo sapiens (Human)		PPI		unmaskedSW		2-115		147-289		76		87.7%				
1	Red Seta View	SGH14009	CA0521921	NC37		Homo sapiens		PPI		unmaskedSW		6-69		318-403		76		85.5%				
1	Red Seta View	SGH14009	CA0521921	Se23 protein		Homo sapiens		PPI		unmaskedSW		19.7%		7-141		134-275		74		84.45%		
1	Red Seta View	SGH14009	CA0521921	coch 1		Homo sapiens		PPI		unmaskedSW		202		73-185		384-497		67		80.7%		
1	Red Seta View	SGH14009	CA0521921	polymerase		Homo sapiens		PPI		unmaskedSW		84-185		442-552		55		74.2%				
1	Red Seta View	SGH14009	CA0521921	diap3k20.3 (Copine 1 (dmbr to KIA0638))		Homo sapiens		PPI		unmaskedSW		25%		76-143		353-429		67		72.8%		
1	Red Seta View	SGH14009	CA0521921	Not protein		Homo sapiens		PPI		unmaskedSW		6-145		373-504		56		71.49%				
1	Red Seta View	SGH14009	CA0521921	60 KDA PROTEIN (60 KDA RIBONUCLEOPROTEIN AD) (KRNPs) (SUDREIN SYNDROME TYPE A ANTIGEN (65-A))		Homo sapiens (Human)		PPI		unmaskedSW		6-145		373-504		56		70.1%				
1	Red Seta View	SGH14009	CA0521921	hexokinase 1		Homo sapiens		PPI		unmaskedSW		18.5%		22-122		749-851		75		68.75%		
1	Red Seta View	SGH14009	CA0521921	KIAA1434 protein		Homo sapiens		PPI		unmaskedSW		20.8%		124-188		578-647		87		67.41%		
1	Red Seta View	SGH14009	CA0521921	Sic348 protein		Homo sapiens		PPI		unmaskedSW		6-91		682-801		61		67.41%				
1	Red Seta View	SGH14009	CA0521921	protoactinin gamma A11 short form protein		Homo sapiens		PPI		unmaskedSW		30-185		475-613		66		86.68%				
1	Red Seta View	SGH14009	CA0521921	protoactinin gamma A11		Homo sapiens		PPI		unmaskedSW		30-185		475-643		66		86.42%				
1	Red Seta View	SGH14009	CA0521921	alphaen 31 (alphaen XXXI)		Homo sapiens		PPI		unmaskedSW		10.2%		20-184		123-308		56		93.4%		
1	Red Seta View	SGH14009	CA0521921	hexokinase 1		Homo sapiens		PPI		unmaskedSW		18.5%		22-122		750-852		73		82.08%		
1	Red Seta View	SGH14009	CA0521921	higin beta-7 subunit		Homo sapiens		PPI		unmaskedSW		16.4%		2-55		192-204		67		88.8%		
1	Red Seta View	SGH14009	CA0521921	higin variant beta4E		Homo sapiens		PPI		unmaskedSW		24.3%		2-36		130-166		73		86.22%		

29/40

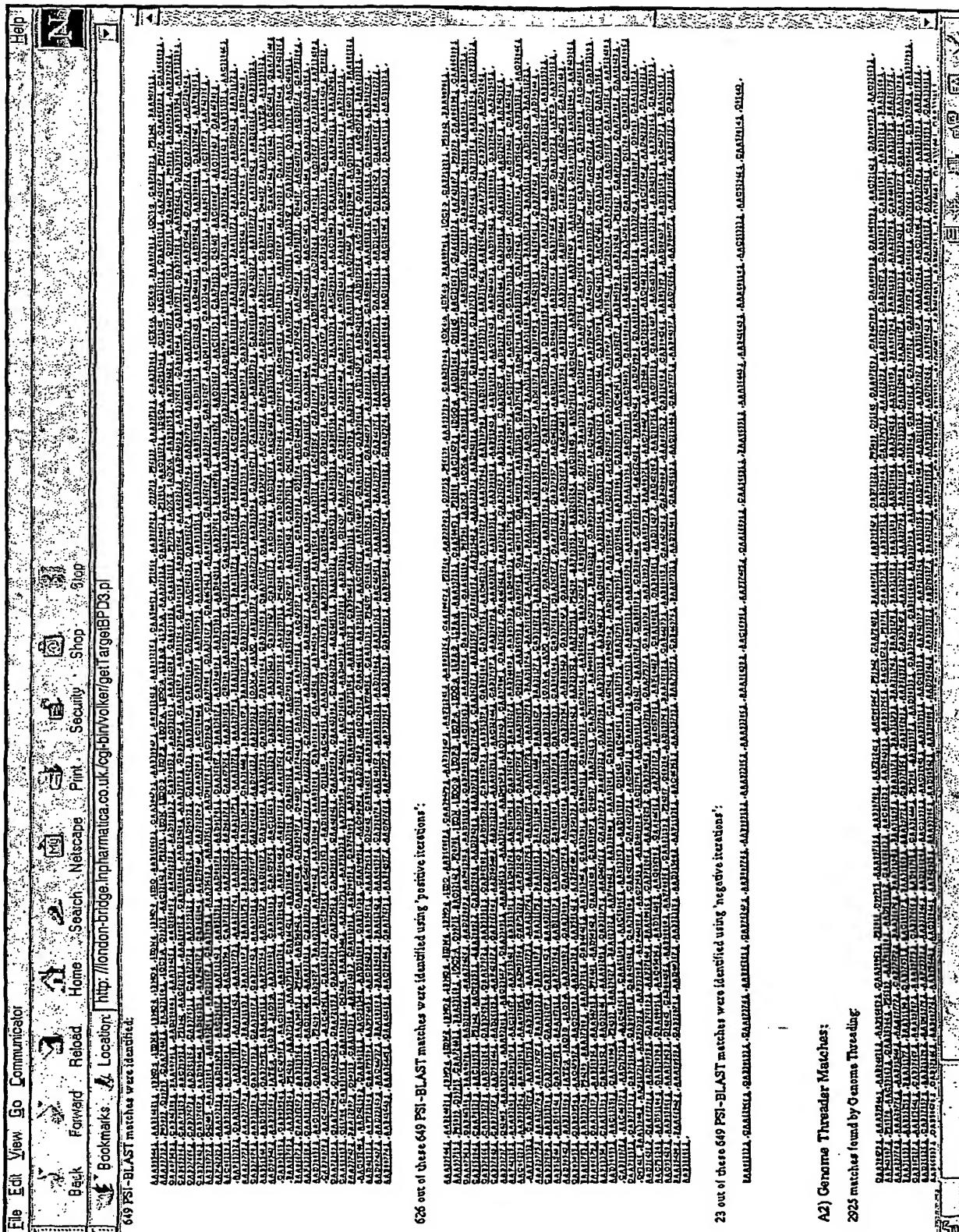


FIG. 18C

SUBSTITUTE SHEET (RULE 26)

30/40

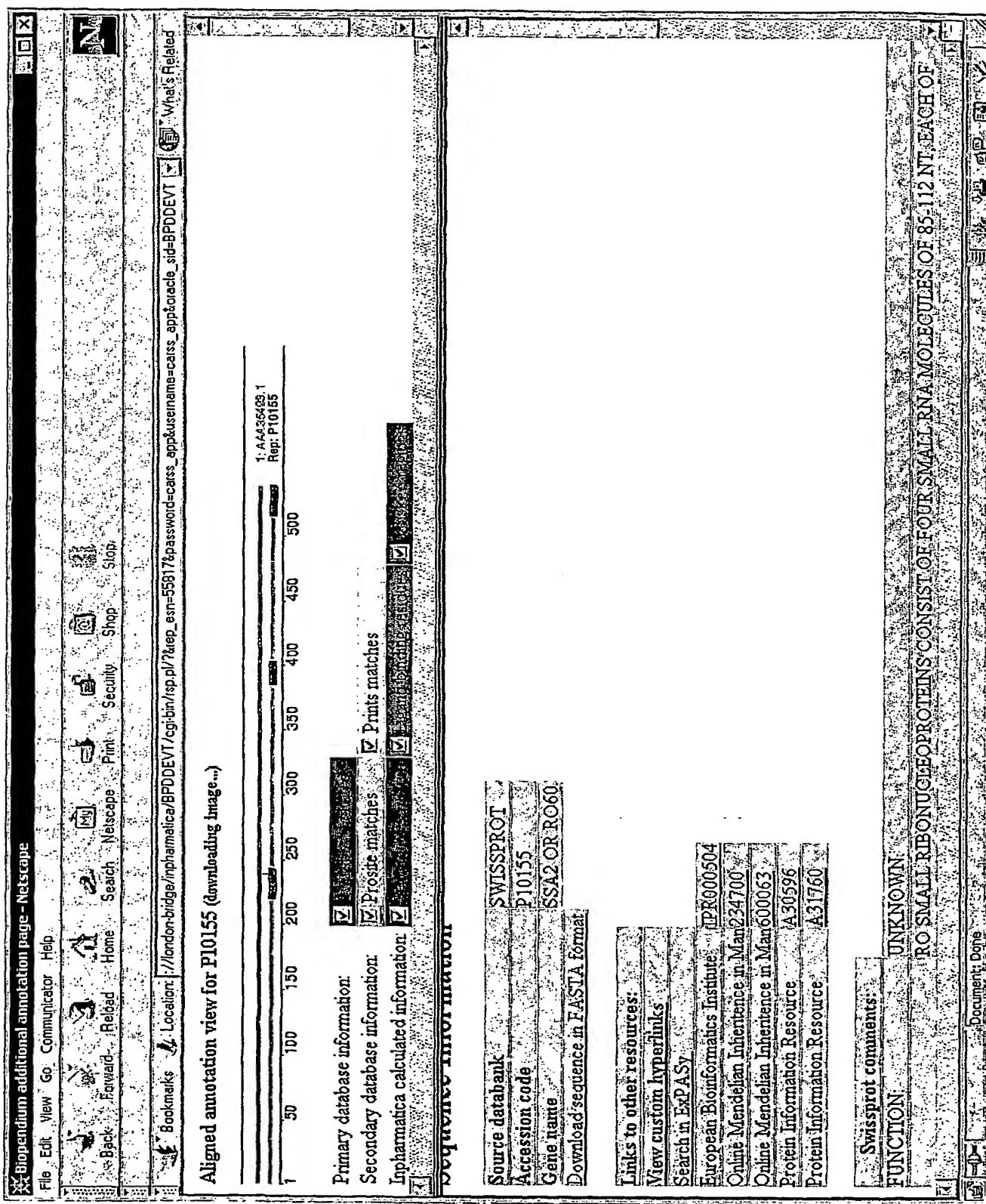


FIG. 19

31/40

FIG. 20

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location: http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi

The Sanger Centre **Pfam**
Protein families database of alignments and HMMs

Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help

Results for gi|133251|sp|P10155|RO60_HUMAN

There were no matches to Pfam-A (including borderline matches) for gi|133251|sp|P10155|RO60_HUMAN

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_8344	1	194	2.3e-103	Align
Pfam-B_10162	195	538	1.8e-165	Align

[538 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: Hypertext linked to swisspfam

Query gi|133251|sp|P10155|RO60_HUMAN/1-194 matching Pfam-B_8344

```

092787 1 MEESVNQMQPLNEKQIANSQDGYYVWQUTDMNRLHRLFCFGSEGGTYYIKE 50
gi|133251|sp|P10155|RO60_HUMAN 1 MEESVNQMQPLNEKQIANSQDGYYVWQUTDMNRLHRLFCFGSEGGTYYIKE 50
092787 51 QKLGLENEREALIRLIEDGR6CEVIQEIKSFSQEGRTTKQEPMLFALAICS 100
gi|133251|sp|P10155|RO60_HUMAN 51 QKLGLENEREALIRLIEDGR6CEVIQEIKSFSQEGRTTKQEPMLFALAICS 100
092787 101 QCSDISTKQARPKAVSEVCRIPTHLPTFIQPKKDLSKESMKCGMWGRALRK 150
gi|133251|sp|P10155|RO60_HUMAN 101 QCSDISTKQARPKAVSEVCRIPTHLPTFIQPKKDLSKESMKCGMWGRALRK 150
092787 151 AIAIDWYNEKCGMALALAVTKYKQRIGWSMKDLLRLSMLKPSSEG 194
gi|133251|sp|P10155|RO60_HUMAN 151 AIAIDWYNEKCGMALALAVTKYKQRIGWSMKDLLRLSMLKPSSEG 194

```

[Align to family](#)

Query gi|133251|sp|P10155|RO60_HUMAN/195-538 matching Pfam-B_10162

```

008848 195 LAIVTKYITKGWKEVHEPEYKEKALSVREKLLKYLEAVEVKVKTKEDELEV 244
gi|133251|sp|P10155|RO60_HUMAN 195 LAIVTKYITKGWKEVHEPEYKEKALSVREKLLKYLEAVEVKVKTKEDELEV 244
008848 245 IMLIEENQLVREHLLTMLKSKEVWWKALLQEMPLTALLRNLGKTRNSVL 294
THITTEWVLUPEMLIITMLKSKEVWWKALLQEMPLTALLRNLGKTRNSVL
008848 295 EPGNSEVSLICEKLSHEKLLKKARLMPFHLLIALETY+TRAGHGLRKLKWI 344
gi|133251|sp|P10155|RO60_HUMAN 295 EPGNSEVSLICEKLSHEKLLKKARLMPFHLLIALETY+TRAGHGLRKLKWI 344
008848 345 PDKDILQRLDRAFTTFTKVEPT6KRFLLAVDVSRSMNRALGSVILNST 394

```

32/40

FIG. 21

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location: http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=Protein&list_uids=9606

LOCUS R060_HUMAN 538 aa **PRI** 01-FEB-1996
DEFINITION 60 KD Ro PROTEIN (60 KD RIBONUCLEOPROTEIN Ro) (R0RNP) (Sjogren SYNDROME TYPE A ANTIGEN (SS-A)).
ACCESSION P10155
PID g133251
VERSION P10155 GI:133251
DBSOURCE swissprot: locus R060_HUMAN, accession P10155, class: standard.
 created: Mar 1, 1989.
 sequence updated: Mar 1, 1989.
 annotation updated: Feb 1, 1996.
 xrefs: gi: gi: 177783, gi: gi: 387656, gi: gi: 387657, gi: gi: 86722, gi: gi: 107626
 xrefs (non-sequence databases): MIM 600053, MIM 234700, PROSITE PS00030
KEYWORDS Ribonucleoprotein; RNA-binding; systemic lupus erythematosus; antigen.
SOURCE
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Deutscher, S.L., Harley, J.B. and Keene, J.D.
TITLE Molecular analysis of the 60-kDa human Ro ribonucleoprotein
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 85 (24), 9479-9483 (1988)
MEDLINE 89071722
REMARK SEQUENCE FROM N.R
REFERENCE
AUTHORS Ben-Chetrit, E., Gandy, B.J., Tan, E.M. and Sullivan, K.F.
TITLE Isolation and characterization of a cDNA clone encoding the 60-kD component of the human SS-A/Ro ribonucleoprotein autoantigen
JOURNAL J. Clin. Invest. 83 (4), 1284-1292 (1989)
MEDLINE 89198084
REMARK SEQUENCE FROM N.R
COMMENT
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. The original entry is available from <http://www.expasy.ch/sprot> and <http://www.ebi.ac.uk/sprot>

[FUNCTION] UNKNOWN.
 [SUBUNIT] Ro SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT, EACH OF WHICH IS COMPLEXED WITH A 60 KD PROTEIN. Ro RNPs MAY ALSO CONTAIN AN ADDITIONAL 52 KD PROTEIN.
 [SUBCELLULAR LOCATION] CYTOPLASMIC.
 [DISEASE] SERA FROM PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS OFTEN CONTAIN ANTIBODIES THAT REACT WITH THE NORMAL CELLULAR Ro PROTEIN AS IF THESE ANTIGEN WAS FOREIGN.
 [SIMILARITY] CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
 [SIMILARITY] STRONG, TO XENOPUS 60 KD Ro PROTEIN.

FEATURES

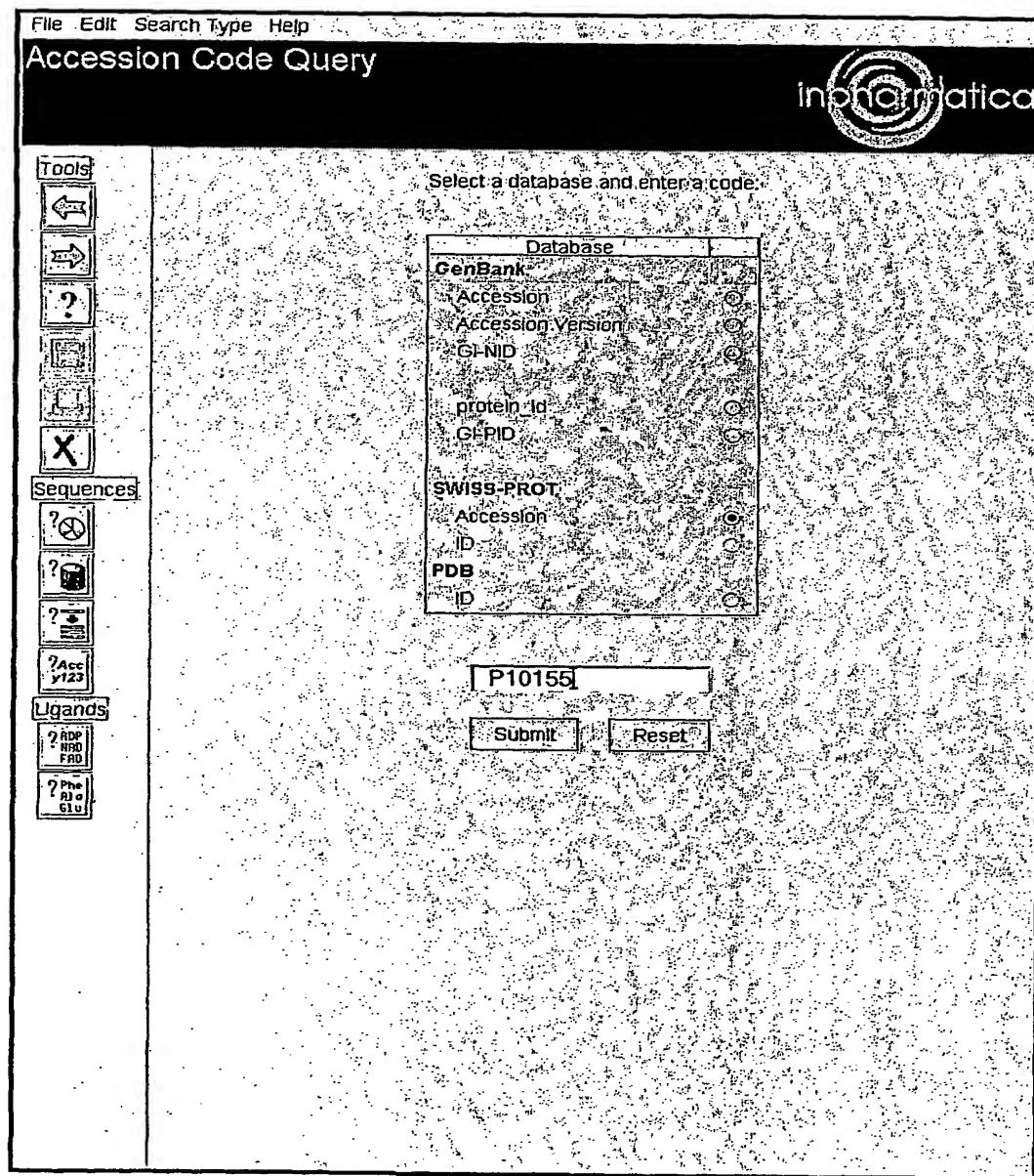
source Location/qualifiers
 1..538
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..538
Protein 1..538
 /product="60 KD Ro PROTEIN"
Region 93..98
 /region_name="Domain"
 /note="RNA-BINDING (RNP2) (BY SIMILARITY)."
Region 124..131
 /region_name="Domain"
 /note="RNA-BINDING (RNP1) (BY SIMILARITY)."
Region 239
 /region_name="Conflict"
 /note="K -> R (IN REF. 2)."
Region 515..538
 /region_name="Conflict"
 /note="GMLDMCCFDTGALDVIRNPTLDMI -> ALQNTILNKSP (IN REF. 2)."
ORIGIN

```

1 meesvnqmqqp lnekqiansq dgywwqvtdm nr1hrflefq seggtyyike qklglennaea
 61 lirlledggrg cevigeiksf sgegrttkqe pmifalaics qesdistkqa afkavsevcr
 121 ipthlftfq fikkdllesmk cgmwvgralrk aiadwymekg gmalalavth vqngwshk
 181 dllrlshlkp sseglaivtk ptkgwkevh elykekalsv etekllkyle avekvkrtkd
 241 elevihliee hrlvrehllt nhikskewk allqemplta llrnlgknta nsylepgnse
 301 vslveeklon ekllrlkaris mnqrvlgsil nastvaaamc lkwrpdeeil kldaaafpk
 361 fktveptgrr fllavdvas mnqrvlgsil nastvaaamc mvvtirtekd s yvvaafsdemv
 421 pcpvttdmrl qqvlaamsqi paggtdcslp miwqktntp advfivitdn etfaggvhpa
 481 ialreyrkkm dipaklivcg ntsngftiad pdrgmldmc gfdtgaldvi mftldmi
  //
```

33/40

FIG. 22A



34/40

Inpharmatica

Query details:
P1015: 60 kDa RIBONUCLEOPROTEIN (60 kDa RIBONUCLEOPROTEIN RO) (RORNP) (SJØRGEN SYNDROME TYPE A ANTIGEN (SS-A)).

Total selected: 1. Page 1 of 2. Total hits: 31. Selected: 1.

View alignment Select all Deselect all

Inpharmatica Genome Threader results.

Cluster	Detail	Code	Title	Organism	%ID	Query len	Target len	Ah score	Method	Confidence
o	D	1ASA:A	CRYO-CRYSTALLOGRAPHY OF A TRUE SU(MN2+ BO... -1JLM	<i>Syphromyrmex</i>	10.9	42 - 172	120 - 250	54	Local	Low (76%)
-	D	1BH0:1	MAC-1 DOMAIN MAGNESIUM COMPLEX	<i>H.sapiens</i>	11.1	373 - 504	6 - 145	56	Local	Low (70%)
-	D	1BH0:2	MAC-1 DOMAIN MAGNESIUM COMPLEX	<i>H.sapiens</i>	11.7	373 - 504	6 - 145	55	Local	Low (67%)
-	D	1DNI:2	MAC-1 DOMAIN METAL-FREE	<i>H.sapiens</i>	11.7	373 - 504	6 - 145	55	Local	Low (67%)
-	D	1BH0:2	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H.sapiens</i>	11.7	373 - 504	6 - 145	55	Local	Low (67%)
-	D	1BHQ:1	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H.sapiens</i>	11.7	373 - 504	6 - 145	55	Local	Low (67%)
-	D	1IDN:1	MAC-1 DOMAIN METAL-FREE	<i>H.sapiens</i>	11.7	373 - 504	6 - 145	55	Local	Low (67%)
-	D	1BEU:2	TRP SYNTHASE (D60N-IPF3 SER) WITH K+ -1DO:1	<i>Syphromyrmex</i>	10.3	24 - 172	102 - 248	53	Local	Low (65%)
-	D	1DO:1	I-DOMAIN FROM INTEGRIN CR3, MG2+ BO... -2TYS:A	<i>H.sapiens</i>	10.9	373 - 489	6 - 131	58	Local	Low (55%)
-	D	1AOX:1	CRYSTAL STRUCTURE OF MUTANT BETA1... -1AOX:1 DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>H.sapiens</i>	11.5	25 - 174	103 - 252	56	Local	Low (53%)
-	D	1UBS:1	TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) WIT... -1BKS:1 TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) FRO...	<i>Syphromyrmex</i>	10.0	41 - 504	45 - 149	53	Local	Low (53%)
-	D	1QGC:1	STRUCTURE OF THE COMPLEX OF AN FAB... -1FM0:1 FOOT AND MOUTH DISEASE VIRUSTYPE...	<i>F.disease</i>	11.5	42 - 174	120 - 234	65	Local	Low (48%)
-	D	1A0K:1	CRYSTAL STRUCTURE OF THE VON WILLE...	<i>H.sapiens</i>	10.9	365 - 504	120 - 251	61	Local	Low (35%)
-	D	1AOX:8	1 DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>H.sapiens</i>	10.3	41 - 504	4 - 150	49	Local	Low (34%)
-	D	1QGC:1	STRUCTURE OF THE COMPLEX OF AN FAB...	<i>F.disease</i>	21.1	277 - 385	10 - 100	62	Local	Low (23%)
-	D	1FM0:1	FOOT AND MOUTH DISEASE VIRUSTYPE...	<i>F.disease</i>	21.1	277 - 385	10 - 100	62	Local	Low (23%)
-	D	1AQ:1	AT DOMAIN OF VON WILLEMANS FACTOR	<i>H.sapiens</i>	7.0	373 - 504	20 - 159	43	Local	Low (21%)
-	D	1BWCB:1	CRYSTAL STRUCTURE OF TYROSINE AMIN...	<i>F.tatoo</i>	5.5	453 - 535	178 - 248	57	Local	Low (15%)
-	D	1TPA:1	TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) IN T...	<i>Syphromyrmex</i>	10.4	44 - 172	122 - 251	51	Local	Low (15%)
-	D	1TQG:1	TRYPTOPHAN SYNTHASE (E.C.4.2.1.20) IN T...	<i>Syphromyrmex</i>	10.4	44 - 172	122 - 251	51	Local	Low (15%)
-	D	1ART:1	ASPARTATE AMINO TRANSFERASE (E.C.26... -1QIT:1	<i>E.coli</i>	10.4	405 - 537	116 - 248	52	Local	Low (15%)
-	D	1ART:2	ASPARTATE AMINO TRANSFERASE FROM...	<i>N.A.</i>	10.5	405 - 537	116 - 248	45	Local	Low (12%)
-	D	1EEFA:1	CRYSTAL STRUCTURE OF YEAST KARYOP...	<i>S.cerevisiae</i>	16.5	35 - 130	298 - 386	49	Local	Low (10%)
-	D	1EEFB:1	CRYSTAL STRUCTURE OF YEAST KARYOP...	<i>S.cerevisiae</i>	15.5	35 - 130	298 - 388	49	Local	Low (10%)
-	D	1ARH:1	ASPARTATE AMINO TRANSFERASE Y225F...	<i>E.coli</i>	10.5	405 - 537	116 - 248	46	Local	Low (10%)
-	D	1ARH:2	ASPARTATE AMINO TRANSFERASE Y225F...	<i>E.coli</i>	10.5	405 - 537	116 - 248	46	Local	Low (10%)
-	D	1EFT:1	CRYSTAL STRUCTURE OF THE MOESIN FE...	<i>N.A.</i>	18.8	82 - 171	121 - 188	55	Local	Low (10%)
-	D	1EFT:2	CRYSTAL STRUCTURE OF THE MOESIN FE...	<i>N.A.</i>	18.8	92 - 171	121 - 188	55	Local	Low (10%)

Sequences

File Edit Search Type Help Tools

► GenBank ► SWISS-PROT ► PDB

FIG. 22B

35/40

FIG. 22B Aligned Sequence Display

Query details:
P10155: 60 kDa Rb protein (60 kDa ribonucleoprotein (RNP)) (Sjögren syndrome type A antigen (SS-A)).

Tools:

Page 1 of 1. Total hits: 13. Selected: 0.

Reverse Maximised PSI-BLAST results:

Cluster	Detail	Code	File	Organism	%ID	Query len.	Target len.	E-value	
o	o	AAF35532.1	Not given	<i>Haemophilus</i>	99.0	1 - 516	1	0.00E+00	
[E]	o	AAF19049.1	Not given	<i>Marinobacter</i>	90.0	1 - 538	1	0.00E+00	
[E]	o	AAC53142.1	Not given	<i>Marusciellus</i>	90.0	1 - 538	1	0.00E+00	
o	o	AAC38601.1	ribonucleoprotein	<i>Xylella</i>	76.0	1 - 538	1	0.00E+00	
o	o	ABB01552.1	Not given	<i>Haemophilus</i>	100.0	1 - 199	1	1.00E-111	
o	o	CAA98241.1	Not given	<i>Celebensis</i>	36.0	7 - 538	54	6.01E+00	
o	o	AAF10833.1	Not given	<i>Brachyellus</i>	34.0	7 - 538	15	5.51E+00	
o	o	AAF46876.1	Not given	<i>Dmetenogaster</i>	17.0	1 - 537	64	6.69E-126	
o	o	CAB46406.1	hypothetical protein	<i>Sphaerotilus</i>	26.0	266 - 381	273 - 372	3	6.00E-25
o	o	AAC55043.1	telomerase protein-1	<i>Marinobacter</i>	21.0	264 - 379	513 - 627	3	6.00E-05
o	o	AB51690.1	telomerase protein component 1	<i>Anoxygenic</i>	23.0	264 - 375	517 - 627	3	2.00E-04
o	o	AAC46601.1	telomerase component p80	<i>Tetrahymenophila</i>	27.0	264 - 328	361 - 425	3	7.00E-04
o	o	AC51107.1	telomerase-associated protein TP-1	<i>Haemophilus</i>	25.0	264 - 327	507 - 569	3	0.00E-03

Sequences:

Organisms:

Proteins:

PPB: PDB: SWISS-PROT: PDB: GenBank:

FIG. 23
AIEye output (January 2, 2001 1:29 PM)

36/40

P10155 MEESYNQHQPLNEKQIANSQDQYVWQVTDNNRLHRLFCFGSEGTYYIKEQKLGLENAAELIRLIEDGRGCEVIEQIKSF
 IJW

P10155 SQEGRTTKQEPHLPALAIQSQCSDISTKQAAFKAVSEVCHIPTHLFTFQPKKDLKESMKCGMWGRALRKAIADWYNEKG
 IJW

P10155 GMALALAVTKYKQANGWSHKDLRLSHLKPSSSEGGLAIYTKEYITKGWKEVHELYKEKALSVETEKLLKYLEAVEKYKRTKD
 IJW

P10155 ELEVINTLICHRLVREHLLTNHLKSGKEVNKAALLQEMPLTLLNLGKMTANSVLEPGNSEVSLYCEKLCNEKLKKARIH
 IJW

P10155 PFHILIALETYKTHGHLAGKLIKWRPDEEILKALDAAFYKTFKTVPTGKHPFL--L--AVDVSASMNQARVLSGSLINASTVA
 IJW

P10155 A---ANCMVYUTATEKDSYVVAFSDEWVPCPVTTDNTLQQVLNANSQI-PAGGTDCSLPNIWAQKNT---PADVFIYF
 IJW

P10155 NEQLKKSKTFLSINQYSEEFRIHFTKEFQNNPNPRLVYKPTTQLLGRATHATGIRKVVRELFINTNGARKNAFKILVW
 IJW

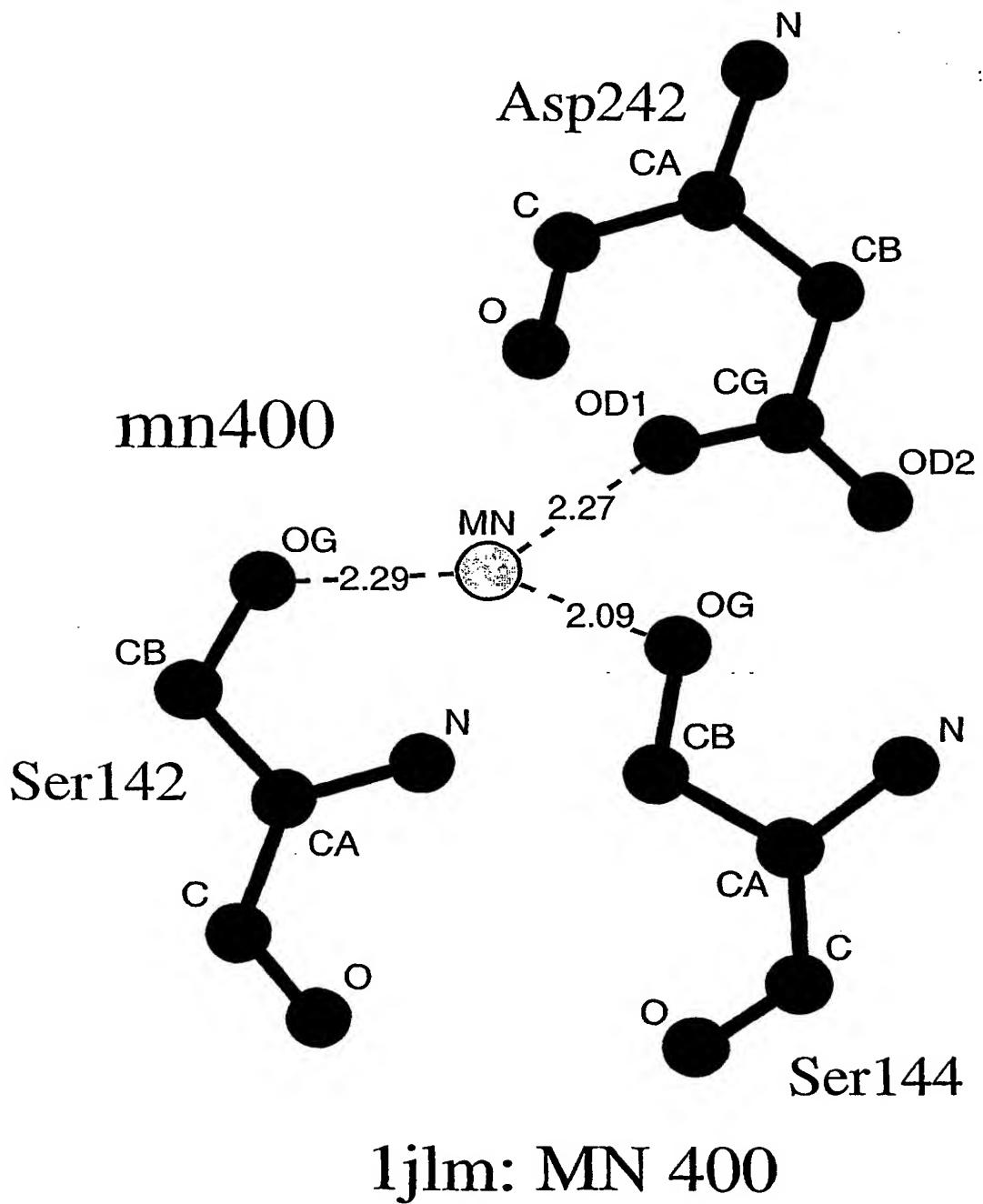
P10155 TDNETFAGGYHHPAIALREYRKKMDIPAKLIVCGMTSNGF-----TIADPDDRG-MLDNGCGFDTGALDVIANFTLDMI
 IJW

P10155 TDGEKFGDPLGYEDVIE-ADEGVIRYVIGVGDAFRSEKSRQELENTIASKPPRHOHVYFQVNNFE--ALKTIQNQLAREKIF
 IJW

P10155 IJW A

37/40

FIG. 24A



38/40

FIG. 24B



FIG. 25

AIEye output (January 4, 2001 3:18 PM)

39/40

P10155 AAF19048.1 AAC38001.1 CAA98241.1	MEESVN-----Q----- MEGSAN-----Q----- MEATMD-----Q----- -----MADELNEFQEAGNFNEEALMRLSNVCAIRRQMLQESDVEITVVDGELKRVPRQMEVKDQVENNAGGFVFP	10 20 30 40 50 60 70
P10155 AAF19048.1 AAC38001.1 CAA98241.1	VTDMNRLHARFLCFFGSEGGTTYIKEOKLGLNEAALIARLIEDGRGC&EVIQEIKSFSOEGRRTKQEPMLFALAICSQCSO VTDMN-VRRFLCFFGSEGGTTYIKEOKLGLNEAALIARLIEDGRGC&EVIQEIKSFSOEGRATAKQEPLLFALAVCSOCAD VSDMNRLRRFLCFFGSEGGTTYIEKKLGQENAEALLRLIEDGKGCEVQIEKTFSQEGRAAKQEPTLFALAVCSQCSO VSDDETQVARRFLILGSDKGSYHQSSEKITYIDNAQRIKIEQNGQHMVLKELALINAENRNPKMNAMIFTLAIARIST	80 90 100 110 120 130 140 150
P10155 AAF19048.1 AAC38001.1 CAA98241.1	ISTKQ----- INTKQ----- IKTKQ----- HDTTKKTECPMLNAYSDYIRALHDSDLIPEVCRATPTHLFEDYDCQT	160 170 180 190 200 210 220 230
P10155 AAF19048.1 AAC38001.1 CAA98241.1	AAFKAVSEVCRIPTHLFTFIQFKKKDLKESMKCGM----- AAFKAVPEVCRIPTHLFTFIQFKKKDLKESMKCGM----- AAFRAVPEVCRIPTHLFTFIQFKKKDLKEGMKCGM----- HDTTEQLFVYDYCQT	240 250 260 270 280 290 300 310
P10155 AAF19048.1 AAC38001.1 CAA98241.1	EKGGMALALAVTKYKORNGWSHKDLRLSH----- EKGGMVALVVTKYKORNGWSHKDLRLSH----- TKDALNAMAVTKYKORNGWSHKDLRLSH----- TTTEKLAAMLTKYPQREGWSHRDLFRLAHPNLMDSRSHQQSEDRLEREQFLRFRAVKGDLVVKAKRMSVEVAEVEKV	320 330 340 350 360 370 380 390
P10155 AAF19048.1 AAC38001.1 CAA98241.1	YKEKALSV----- YKEKALSV----- YKEKELSP----- WDKKALKLPYTHEQSLIKEQSRALNLVEAYLKLKNEQSEEVIVAAIKHHGLVREHLPTTSLN SKLVWETLFDVSMPMT	400 410 420 430 440 450 460

40/40

FIG. 25 (contd.)
AIEye output (January 4, 2001 3:18 PM)

470 Y KTFKTVEPPTGKRFLLA VD VSASMNQ RVLGS I L NASTVAAACMV VVTRATEKDSYVVA FSD EMVPCPVTTDMTLQQVLM
 P10155 YTTFKTVEPPTGKRFLLA VD VSASMNQ RALGS VLNASTVAAACMV VVTRATEKESSVVA FACDMVPPVTTDMTLQQVLT
 AAF19049.1 AAC38001.1 CAA98241.1
 480 YKSFKLVEPTGKRFLLA ID VSASMNQ RVLGS I L NASVVA AAMCML V A R T E K D S H M V A F S D E M L P C P I T V N M L L H E V V E
 YKAFVNAPPTGKRYCLAD VSGSM T S R V S S P L S C R E A T G M S L I N L H N E A E V R C V A F C D K I T E L P F T K D W K I G Q V N D
 490 YKA FVNAPPTGKRYCLAD VSGSM T S R V S S P L S C R E A T G M S L I N L H N E A E V R C V A F C D K I T E L P F T K D W K I G Q V N D
 500 YKA FVNAPPTGKRYCLAD VSGSM T S R V S S P L S C R E A T G M S L I N L H N E A E V R C V A F C D K I T E L P F T K D W K I G Q V N D
 510 YKA FVNAPPTGKRYCLAD VSGSM T S R V S S P L S C R E A T G M S L I N L H N E A E V R C V A F C D K I T E L P F T K D W K I G Q V N D
 520 YKA FVNAPPTGKRYCLAD VSGSM T S R V S S P L S C R E A T G M S L I N L H N E A E V R C V A F C D K I T E L P F T K D W K I G Q V N D
 530 YKA FVNAPPTGKRYCLAD VSGSM T S R V S S P L S C R E A T G M S L I N L H N E A E V R C V A F C D K I T E L P F T K D W K I G Q V N D
 540 YKA FVNAPPTGKRYCLAD VSGSM T S R V S S P L S C R E A T G M S L I N L H N E A E V R C V A F C D K I T E L P F T K D W K I G Q V N D
 550 A M S Q I P A G G T D C S L P M I W A Q K T N T P A D V F I V F T D N E T F A G Q V H P A I A L R E Y R K K M D I - P A K L I V C G M T S N G F T I A D P D
 P10155 A M N K V P A G G T D C S L P M I W A Q K T G T A A D V F I V F T D N E T F A G Q V H P A V A L R E Y R K K M D I - P A K L I V C G M T S N G F T I A D P D
 AAF19049.1 AAC38001.1 CAA98241.1
 560 YVN NLD F G R T D C G L P M T W A T E N N L K F D V F I I Y T D N D T W A G E I H P F E A I K K Y R E A S Q I H D A K V I V M A M Q A Y D Y S I A D P S
 570 YVN NLD F G R T D C G L P M T W A T E N N L K F D V F I I Y T D N D T W A G E I H P F E A I K K Y R E A S Q I H D A K V I V M A M Q A Y D Y S I A D P S
 580 YVN NLD F G R T D C G L P M T W A T E N N L K F D V F I I Y T D N D T W A G E I H P F E A I K K Y R E A S Q I H D A K V I V M A M Q A Y D Y S I A D P S
 590 YVN NLD F G R T D C G L P M T W A T E N N L K F D V F I I Y T D N D T W A G E I H P F E A I K K Y R E A S Q I H D A K V I V M A M Q A Y D Y S I A D P S
 600 YVN NLD F G R T D C G L P M T W A T E N N L K F D V F I I Y T D N D T W A G E I H P F E A I K K Y R E A S Q I H D A K V I V M A M Q A Y D Y S I A D P S
 610 YVN NLD F G R T D C G L P M T W A T E N N L K F D V F I I Y T D N D T W A G E I H P F E A I K K Y R E A S Q I H D A K V I V M A M Q A Y D Y S I A D P S
 620 YVN NLD F G R T D C G L P M T W A T E N N L K F D V F I I Y T D N D T W A G E I H P F E A I K K Y R E A S Q I H D A K V I V M A M Q A Y D Y S I A D P S
 630 D R G M L D M C G F D T G A L D V I R N F T L D M I
 P10155 D R G M L D M C G F D T A A L D V I R N F T L D V I
 AAF19049.1 AAC38001.1 CAA98241.1
 640 D R G M L D I C G F D S G A L D V I R N F T L D L I
 650 D A G M L D I T G F D S A V P Q I V H E F V T G K I